Associating adverse drug effects with protein targets by integrating adverse event, *in vitro* bioactivity, and pharmacokinetic data

Ines Arlette Smit

Darwin College

July 2020

This thesis is submitted for the degree of Doctor of Philosophy

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text.

It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.

It does not exceed the prescribed word limit for the Physics and Chemistry Degree Committee.

Summary

Adverse drug effects are unintended and undesirable effects of medicines, causing attrition of molecules in drug development and harm to patients. To anticipate potential adverse effects early, drug candidates are commonly screened for pharmacological activity against a panel of protein targets. However, there is a lack of large-scale, quantitative information on the links between routinely screened proteins and the reporting of adverse events (AEs). This work describes a systematic analysis of associations between AEs observed in humans and bioactivities of drugs while taking into account drug plasma concentrations.

In the first chapter, post-marketing drug-AE associations are derived from the United States Food and Drug Administration Adverse Event Reporting System using disproportionality methods, while applying Propensity Score Matching to reduce confounding factors. The resulting drug-AE associations are compared to those from the Side Effect Resource, which are primarily derived from clinical trials. The analysis reveals that the datasets generally share less than 10% of reported AEs for the same drug and have different distributions of AEs across System Organ Classes (SOCs).

Using the drugs from the two AE datasets described in the first chapter, the second chapter integrates corresponding bioactivities, i.e. measured potencies and affinities from the ChEMBL database and ligand-based target predictions obtained with the tool PIDGIN, with drug plasma concentrations compiled from literature, such as C_{max}. Compared to a constant bioactivity cut-off of 1 μM, using the ratio of the unbound drug plasma concentration over the drug potency, i.e. C_{max}/XC_{50}, results in different binary activity calls for protein targets. Whether deriving activity calls in this way results in the selection of targets with greater relevance to human AEs is investigated in the third chapter, which computes relationships between targets and AEs using different measures of statistical association. Using the C_{max}/XC_{50} ratio results in higher Likelihood Ratios and Positive Predictive Values (PPVs) for target-AE associations that were previously reported in the context of secondary pharmacology screening, at the cost of a lower recall, possibly due to the smaller size of the dataset with available plasma concentrations. Furthermore, a large-scale quantitative assessment of bioactivities as indicators of AEs reveals a trade-off between the PPV and how many AE-associated drugs can potentially be detected from *in vitro* screening, although using combinations of targets can improve the detection rate in ~40% of cases at limited cost to the PPV. The work highlights AEs most strongly related to bioactivities and their SOC distribution.

Overall, this thesis contributes to knowledge of the relationships between *in vitro* bioactivities and empirical evidence of AEs in humans. The results can inform the selection of proteins for secondary pharmacology screening and the development of computational models to predict AEs.

Acknowledgements

I would like to express my sincere gratitude to:

My academic supervisor Dr. Andreas Bender for giving me the opportunity to undertake this project, for his guidance and supervision, for providing opportunities for networking and presenting, and for always keeping in mind the practical relevance of scientific work.

My industrial supervisor Dr. Thierry Hanser for his guidance, ideas, insightful discussions, and for welcoming me at Lhasa Limited in Leeds.

My former and current colleagues in the Bender group. Dr. Avid Afzal for welcoming me in the group, help with database installations for this project and discussing statistics. Dr. Kathryn Giblin for sharing her insights into the analysis of adverse event data and resources related to the FAERS database, which helped me get started with this work, and her kind support. Chad Allen for his help with curating the plasma concentrations. Dr. Fredrik Svensson, Dr. Samar Mahmoud, Dr. Azedine Zoufir, and Peter Wright for help with structuring my work, helpful scientific discussions, proofreading, and encouragement. Dr. Lewis Mervin for help with the target prediction.

At the Department of Chemistry: Computer Officers Dr. Catherine Pitt and Dr. Owen Johnson for keeping the server running, and Susan Begg for administrative assistance.

Lhasa Limited for funding this research.

The British Toxicology Society for the opportunity to present my work at the 2019 congress.

The organisations providing publicly available data and software used in the work.

My friends Avid and Fredrik, whose encouragement, humour and wise words made a world of difference to me, and Chad, Peter, and Martina for whose support I am very grateful.

Marc, for his loving support and sense of perspective.

List of publications

Svensson F, Zoufir A, Mahmoud S, Afzal AM, **Smit I**, Giblin KA, Clements PJ, Mettetal JT, Pointon A, Harvey JS, Greene N, Williams RV, Bender A. Information-Derived Mechanistic Hypotheses for Structural Cardiotoxicity. Chem Res Toxicol. 2018 Nov 19;31(11):1119-1127.

This work is reviewed as part of the literature review in the Introduction.

Smit IA, Afzal AM, Allen CHG, Svensson F, Hanser T, Bender A. Systematic analysis of protein targets associated with adverse events of drugs from clinical trials and post-marketing reports. bioRxiv 2020.06.12.135939; doi: https://doi.org/10.1101/2020.06.12.135939

Additional Data Files related to this work are available from the University of Cambridge repository at: https://doi.org/10.17863/CAM.53868

Sections from this preprint are included in chapters 2 to 4. The author performed all analyses and produced all writing and figures while taking into account comments from co-authors.

Table of Contents

List of abbreviations

5-HT	Serotonin
AC_{50}	Concentration required for 50% activity
ACE	Angiotensin-converting enzyme
ADME	Absorption, distribution, metabolism and elimination
ADR	Adverse drug reaction
ADRB1	β-1 adrenergic receptor
AE	Adverse event
ATC	Anatomical Therapeutic Chemical Classification
AUC	Area under the curve
AUROC	Area under the receiver-operator curve
BCPNN	Bayesian confidence propagation neural network
BSEP	Bile salt export pump
CA	Carbonic anhydrase
CA5A	Carbonic anhydrase 5A
CA5B	Carbonic anhydrase 5B
C_{max}	Maximum concentration
COX-1	Cyclooxygenase 1
C_t	Concentration total
C_u	Concentration unbound
EBGM	Empirical Bayes Geometric Mean
EC_{50}	Half-maximal effective concentration
EMA	European Medicines Agency
FAERS	FDA Adverse Event Reporting System
FAERS AEOLUS	FAERS Adverse Event Open Learning through Universal Standardisation (AEOLUS)
FDA	US Food and Drug Administration
FN	False negatives
FP	False positives
f_u	Fraction unbound
GI	Gastrointestinal
GPCR	G protein-coupled receptor

hERG	Human ether-a-go-go-related gene
HLGT	High Level Group Term
HLT	High Level Term
HTTK	High-Throughput Toxicokinetics
IC	Information Component
IC_{50}	Half-maximal inhibitory concentration (Concentration required for 50% inhibition)
IQR	Interquartile range
K_d	Dissociation constant
K_i	Inhibition constant
LLT	Low Level Term
Log	Logarithm
LR	Likelihood Ratio
MedDRA	Medical Dictionary for Regulatory Activities
MEDRT	Medication Reference Terminology
MeSH	Medical Subject Heading
MHRA	Medicines and Healthcare products Regulatory Agency
NPV	Negative Predictive Value
PDE3	Phosphodiesterase 3
PDGFR	Platelet-derived growth factor
PIDGIN	Prediction IncluDinG Inactivity (target prediction tool)
PPB	Plasma protein binding
PPV	Positive Predictive Value
Pr	Probability
PRR	Proportional Reporting Ratio
PSM	Propensity score
PT	Preferred Term
qAOP	Quantitative Adverse Outcome Pathways
QST	Quantitative Systems Toxicology
ROR	Reporting Odds Ratio
SEA	Similarity Ensemble Approach
SIDER	Side Effect Resource
SLC6A2	Norepinephrine transporter
SOC	System Organ Class

STITCH	Search Tool for Interacting Chemicals
TdP	Torsade de pointes
TN	True negatives
TP	True positives
TSPO	Translocation protein
US	United States
VEGRF	Vascular endothelial growth factor receptor
WHO	World Health Organization

1 Introduction

1.1 Adverse drug effects

Medicines can, in addition to their therapeutic effects, have unintended effects, and when these are harmful they are referred to as 'adverse drug effects', defined as "an adverse outcome that can be attributed, with some degree of probability, to an action of a drug" *(1, 2)*. Adverse drug effects are experienced by patients as adverse drug reactions (ADRs), but both terms refer to the same phenomenon *(1)*. For example, the patient information leaflet of diclofenac tablets, a painkiller from the non-steroidal anti-inflammatory drug class, lists possible side effects according to frequency (Table 1.1). Some ADRs occur with the normal doses used for therapy *(3)*, whereas other effects can arise with off-label use, overdose, misuse, abuse, and medication errors *(2, 4)*.

Table 1.1 Examples of possible side effects listed on a patient information leaflet for diclofenac tablets
Adapted from *(5)*.

Frequency	**Example of listed side effect**
Uncommon (may affect up to 1 in 100 people)	Sudden and crushing chest pain (signs of myocardial infarction or heart attack)
Rare (may affect up to 1 in 1,000 people)	Any sign of bleeding in the stomach or intestine
Very rare (may affect up to 1 in 10,000 people)	Fits
Unknown	Chest pain, which can be a sign of a potentially serious allergic reaction called Kounis syndrome

When monitoring patients, a causal link between the drug exposure and adverse outcomes may not be established, thus the more inclusive terms 'adverse event' (AE), or 'adverse experience' are used *(1)*. Therefore, the United States (US) Food and Drug Administration (FDA) distinguishes between an AE, defined as: "any untoward medical occurrence associated with the use of a drug

in humans, whether or not considered drug related", and an adverse reaction, defined as "any adverse event caused by a drug" *(1, 6)*.

ADRs can harm patients by leading to life-threatening injury, prolonged hospital stays, disability and mortality *(7, 8)*. While the exact incidence of ADRs is difficult to determine *(2, 9)*, one study estimated the incidence of serious ADRs among hospitalised patients in the US to be 6.5% *(10)*. The percentage of hospital admissions in Europe due to ADRs ranged from 0.5% to 12.8% in different studies *(8)*, whereas a study in England estimated that 4% of hospital beds are occupied due to ADRs *(11)*. In patients admitted due to ADRs, mortality attributed to the ADR was 2.3% *(11)*. Thus, ADRs account for considerable harm to patients and costs to healthcare systems *(11)*.

Safety issues are also a major obstacle in the development of new drugs. Waring et al. investigated drug candidates developed by four major pharmaceutical companies between 2000 and 2010 that were terminated, i.e. not progressed into further clinical studies *(12)*. Among early candidate compounds, 59% of terminations were due to failure in preclinical toxicology *(12)*. In phase I and phase II clinical trials, 25% of terminations were caused by a lack of clinical safety, making this the leading and second leading causes of attrition in these phases, respectively *(12)*. Similarly, an analysis of project closures in AstraZeneca between 2005 and 2010 showed that 62% of phase I and 35% of phase II projects were terminated due to clinical safety issues *(13)*. These statistics highlight that drug safety and adverse drug effects are a major challenge in drug development and that there is a need for early and accurate prediction of adverse drug effects.

1.1.1 Types of adverse drug reactions

ADRs can broadly be categorised in type A and type B reactions (Table 1.2). In this classification, type A reactions are dose-dependent and predictable from the primary, secondary and safety pharmacology *(1, 14, 15)*. Type B reactions are not dose-related and not predictable from the known pharmacology, and are sometimes referred to as idiosyncratic *(1, 14, 15)*.

Table 1.2 Classification of ADRs

Adapted from *(1, 15–17)*.

Adverse drug reaction	**Description**	**Incidence**
Type A	• Dose-dependent • Predictable from primary, secondary and safety pharmacology • Mechanism related to pharmacology	Relatively frequent (usually >1 in 100 users)
Type B	• Not related to dose • Not related to pharmacology • Not predictable, unexpected • Mechanism often not known	Minority (often <1 in 1000 users)
Type C	• Long-term adaptive changes • Increased frequency of a disease in population • Not predictable, unexpected • Mechanism uncertain	Not reported

Effects arising from the primary pharmacology, i.e. on-target effects, are those due to interaction with the primary therapeutic target of the drug, which is the protein believed to be responsible for the therapeutic effect of the drug *(18, 19)*. An example is hypoglycaemia due to sulphonylurea used in diabetes *(14)*. In contrast, off-target effects arise from unintended interaction with secondary targets, such as QT prolongation caused by inhibition of the human ether-a-go-go-related gene-encoded voltage-dependent potassium channel (hERG) by a variety of drugs, e.g. terfenadine *(19, 20)*. The mediation of cellular effects due to on-target and off-target interactions is illustrated in Figure 1.1. In contrast, an example of a type B reaction, which are not caused by pharmacological interactions, is anaphylactic shock resulting from an allergic reaction to e.g. penicillin *(14, 21)*.

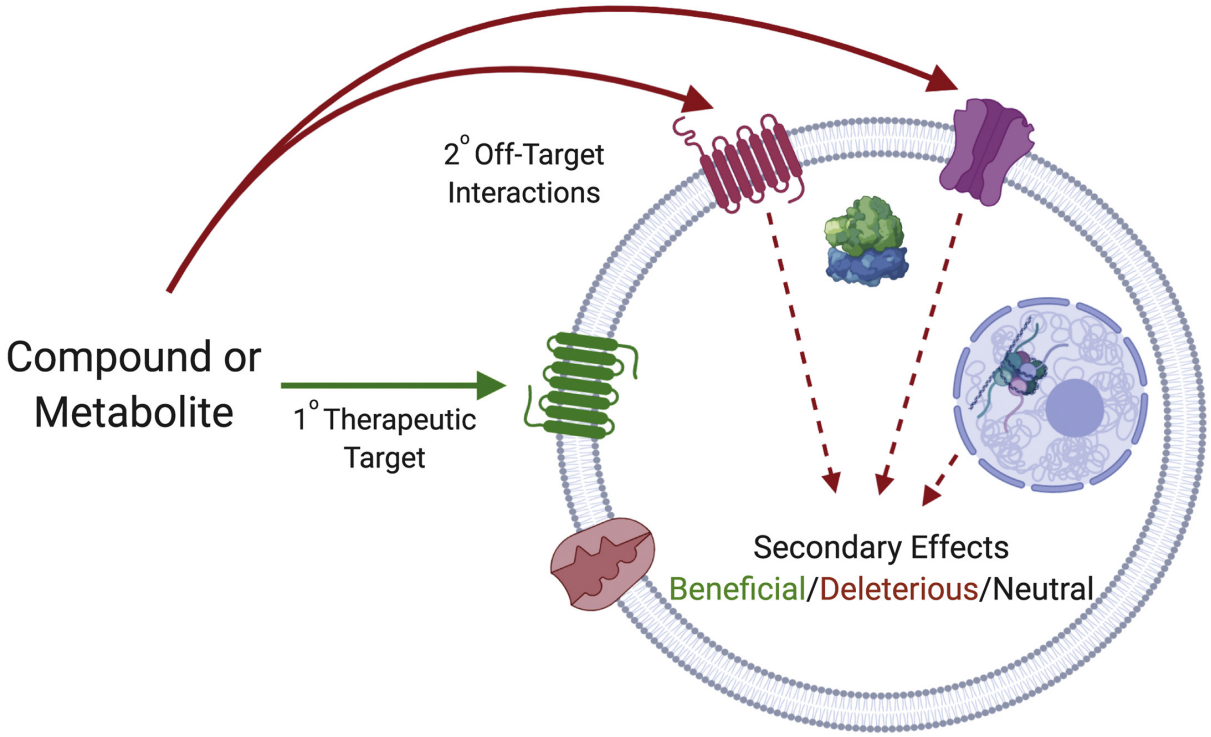

Figure 1.1 Schematic illustration of drugs interacting with primary and secondary pharmacological targets which may result in different effects. Secondary target interactions may cause undesirable effects.

Reprinted from Jenkinson et al. *(22)* with permission from Elsevier.

The A/B classification is imperfect and there exist numerous ADRs that do not match the descriptions of the categories *(1)*. As such, extensions to the classification have been proposed, although the ADR incidence for these classifications is not available *(1)*. For example, the A/B classification was extended with type C reactions (Table 1.2), referring to effects of long-term use, especially an increase in the frequency of a disease that is already relatively common in the population *(1, 16, 17)*. An example is cardiovascular events related to rofecoxib *(17)*. However, the dose-relatedness of type C reactions has not been described, illustrating how the ADRs categories can overlap *(1)*. Another review categorized drug-induced toxicities in five groups: hypersensitivity and related immunological reactions; on-target pharmacology; off-target pharmacology; biological activation to toxic metabolites; and idiosyncratic toxicities, where idiosyncratic effects were regarded as rare occurrences of the former mechanisms *(23)*.

A meta-analysis of the incidence of ADRs in hospitalised patients or leading to hospital admissions reported that ~75% of ADRs were type A reactions and ~25% type B reactions *(10)*. These statistics give a broad idea of the importance of pharmacology in causing ADRs and the potential to predict ADRs from pharmacological interactions *(15, 24)*.

1.1.2 Medical Dictionary for Regulatory Activities (MedDRA) hierarchy

The Medical Dictionary for Regulatory Activities (MedDRA) is the main standardised medical vocabulary currently in use for recording ADRs in regulatory submissions and the study of ADRs *(25)*.

The vocabulary is organised as a hierarchy with five levels. At the highest, most general level are the 27 System Organ Classes (SOCs) (Figure 1.2). The most specific terms are the Lowest Level Terms (LLTs), which can be interpreted as synonyms that are mapped to a Preferred Term *(26)*. A Preferred Term (PT) is defined as "a distinct descriptor (single medical concept) for a symptom, sign, disease diagnosis, therapeutic indication, investigation, surgical or medical procedure, and medical social or family history characteristic" *(27)*. Directly above the PT level is the High Level Term (HLT), grouping related PTs *(27)*. In turn, HLTs are grouped in High Level Group Terms (HLGT) under the SOC. To illustrate this hierarchy, the path from the SOC Cardiac disorders to the PT 'Torsade de pointes' (TdP) is shown in Figure 1.3.

System Organ Classes

Blood and lymphatic system disorders
Cardiac disorders
Congenital, familial and genetic disorders
Ear and labyrinth disorders
Endocrine disorders
Eye disorders
Gastrointestinal disorders
General disorders and administration site conditions
Hepatobiliary disorders
Immune system disorders
Infections and infestations
Injury, poisoning and procedural complications
Investigations
Metabolism and nutrition disorders
Musculoskeletal and connective tissue disorders
Neoplasms benign, malignant and unspecified (incl cysts and polyps)
Nervous system disorders
Pregnancy, puerperium and perinatal conditions
Product issues
Psychiatric disorders
Renal and urinary disorders
Reproductive system and breast disorders
Respiratory, thoracic and mediastinal disorders
Skin and subcutaneous tissue disorders
Social circumstances
Surgical and medical procedures
Vascular disorders

Figure 1.2 The 27 SOCs in the MedDRA (version 21.1). These form the highest, most general level in the hierarchy.

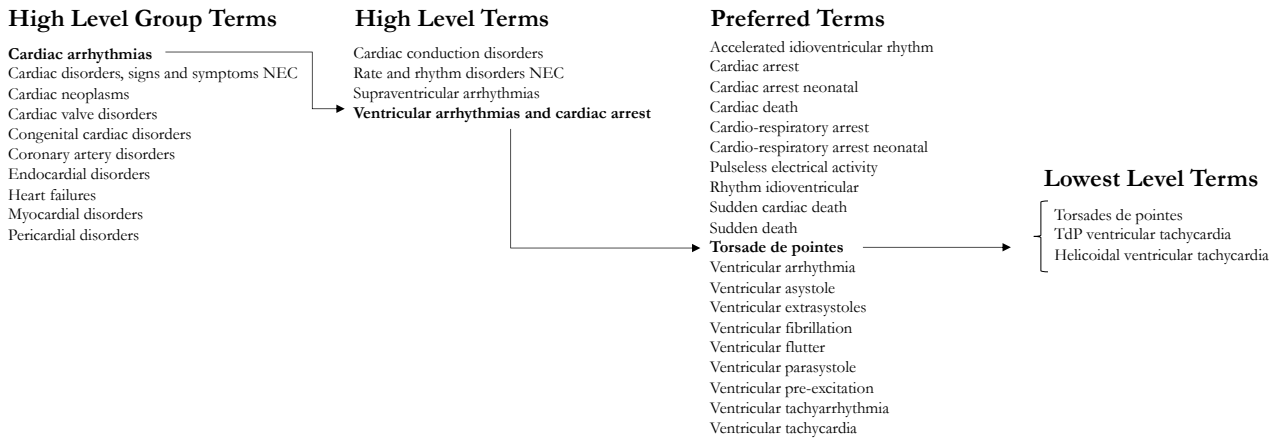

Figure 1.3 The path from the SOC 'Cardiac Disorders' to the PT TdP in the MedDRA hierarchy (version 21.1).

1.2 Pharmacokinetics

In most instances, the effect of drugs is dependent on their concentration at the molecular sites of action, such as a receptor *(28, 29)*. Pharmacokinetics are concerned with (1) the relationship between dose and the unbound drug concentration at the site of action and (2) the drug concentration over time *(29, 30)*.

The concentration at the site of action depends on the dose of the drug administered and on the drug's disposition, which encompasses the processes of absorption, distribution, metabolism and elimination (ADME) *(28, 30)*. These factors combined result in changes of the drug concentration over time *(28, 29)*. Different drugs have different ADME properties, and individuals also differ, for example in their metabolic capacity, which is why the exposure of the tissue or target site is important rather than the dose of the drug *(31, 32)*.

While concentrations at the site of action can typically not be measured, drug plasma concentrations can be measured more easily *(29)*. A generic time course of the drug concentration in plasma is shown in Figure 1.4. The drug concentration in plasma can be measured at different intervals after drug administration to derive exposure parameters such as maximum concentration C_{max} and area under the plasma concentration-time curve (AUC) *(28, 30)*. The C_{max} refers to the maximum concentration reached during a dosing interval *(30)*. The AUC is the total area under the plasma drug concentration time-curve and can be used to understand the systemic exposure to the drug (Figure 1.4) *(30, 31)*.

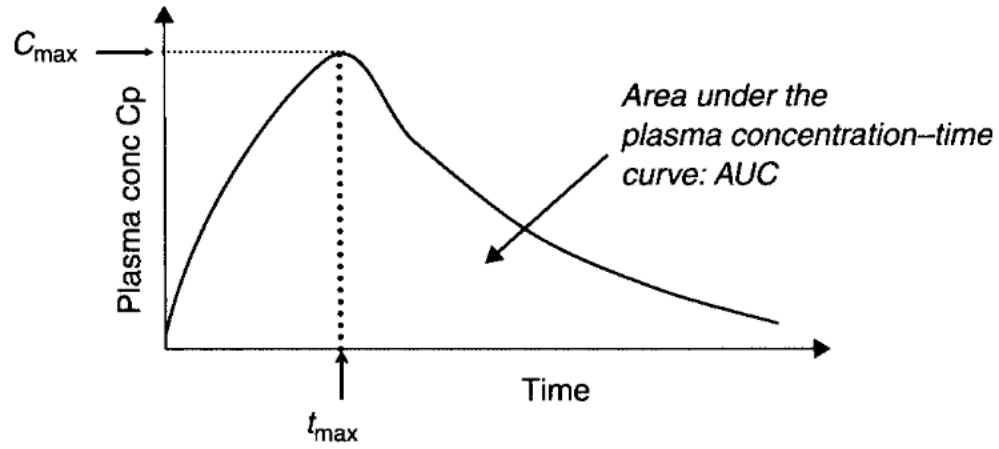

Figure 1.4 Generic plasma concentration-time plot and pharmacokinetic parameters
Reproduced from Rolan and Molnár *(29)* with permission from John Wiley and Sons © 2013.

The plasma concentration can differ from the concentration in various tissues and the brain, but the since the latter cannot be measured non-invasively, it is common to use plasma concentrations as a surrogate measurement *(22, 31, 33)*.

In the blood, drugs reversibly bind to plasma proteins, i.e. plasma protein binding (PPB) and similar processes happen in tissues *in vivo (32)*. Only the unbound (free) drug is available for interaction with proteins and therefore most directly related to a drug's pharmacological effects, although there are some exceptions to this *(31, 32)*.

The extent of PPB varies between drugs, thus to derive the relevant free concentration, this must be taken into account. The unbound concentration is derived by multiplying the total concentration by the free drug fraction:

$$C_u = C_t * f_u \tag{1.1}$$

where C_u is the unbound concentration, C_t is the total concentration, and f_u is the fraction unbound *(30, 32)*.

1.3 Pharmacodynamics and drug action

Pharmacodynamics is concerned with the relationship between the unbound drug concentration at the site of action and the drug response, such as therapeutic or toxic effects *(30)*. Drugs can

bind to different types of proteins, such as enzymes, ion channels, transporters, and receptors, of which receptors are the largest class of drug targets *(18)*.

Compounds that produce similar effects at receptors as endogenous signalling molecules are called agonists *(28, 34)*. Compounds that bind to the receptor but do not produce any direct cellular effect are called antagonists, which can reduce the action of the receptor by preventing the endogenous ligands from binding *(28, 34, 35)*. Agonism at a protein target can have different or opposite effects to antagonism *(19)*.

Agonists that are able to elicit a full functional response as they bind to the target are full agonists, whereas those that are only able to elicit a partial response, even at high concentrations, are partial agonists, as illustrated in Figure 1.5. Thus, partial agonists have lower intrinsic efficacy than full agonists, and antagonists lack efficacy *(35, 36)*. Efficacy is the extent of functional change or cellular effect resulting from drug binding *(28, 34)*

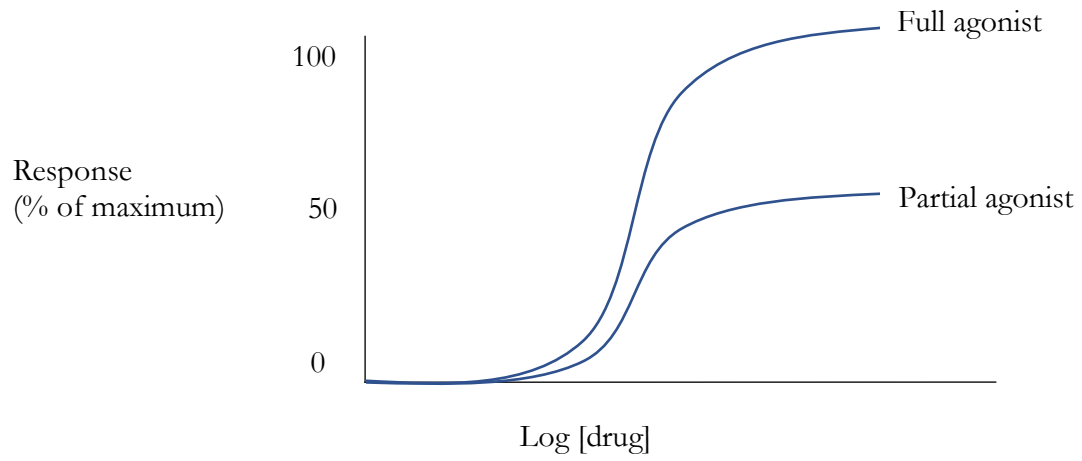

Figure 1.5 Partial agonists have a lower capacity to elicit a cellular response, i.e. lower intrinsic efficacy, than full agonists.

Antagonists that compete with the endogenous ligand for the same binding site are competitive antagonists *(36)*. Inverse agonism is similar to competitive antagonism, but happens when the receptor is constitutively active and the ligand has higher affinity for the inactive state than the active state of the receptor, reducing the overall activity level (Figure 1.6). This is in contrast with neutral antagonists, which have equal affinity to the active and inactive state *(35, 36)*.

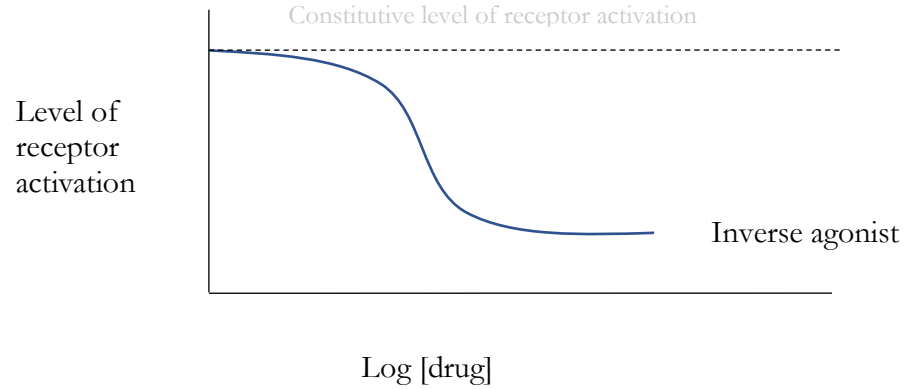

Figure 1.6 Effect on inverse agonists on the level of receptor activation
Adapted from *(35)*.

Allosteric antagonists bind to a separate binding site on the protein, the allosteric binding site, causing a decrease in the affinity of the protein for the endogenous agonist *(36)*. This is referred to as non-competitive antagonism *(36)*. Alternatively, allosteric modulation that potentiates the endogenous agonist's response is achieved by allosteric agonists *(36)*. Another mechanism of functional non-competitive antagonism is due to covalent binding or very slow dissociation from the main binding site *(36)*, although this has also been called irreversible competitive antagonism *(35)*. Lastly, uncompetitive inhibition refers to an inhibitor binding only to the enzyme-substrate complex *(37, 38)*, and sometimes uncompetitive antagonism is used to describe antagonists that can only bind to receptors upon activation of the receptor by the agonist *(38, 39)*.

The bioactivity of drugs can be quantified as their potency, which is the amount of drug required to produce an effect of given intensity *(40)*. Two common measures of potency are the half-maximal effective concentration (EC_{50}/AC_{50}) for agonists and half-maximal inhibitory concentration (IC_{50}) for antagonists or enzyme inhibitors *(37, 40)*. These measures refer to the concentration required to reach 50% of the maximal effect of the compound on a biochemical function, and are derived from the concentration-effect curve (Figure 1.7 and Figure 1.8) *(28, 35, 40)*.

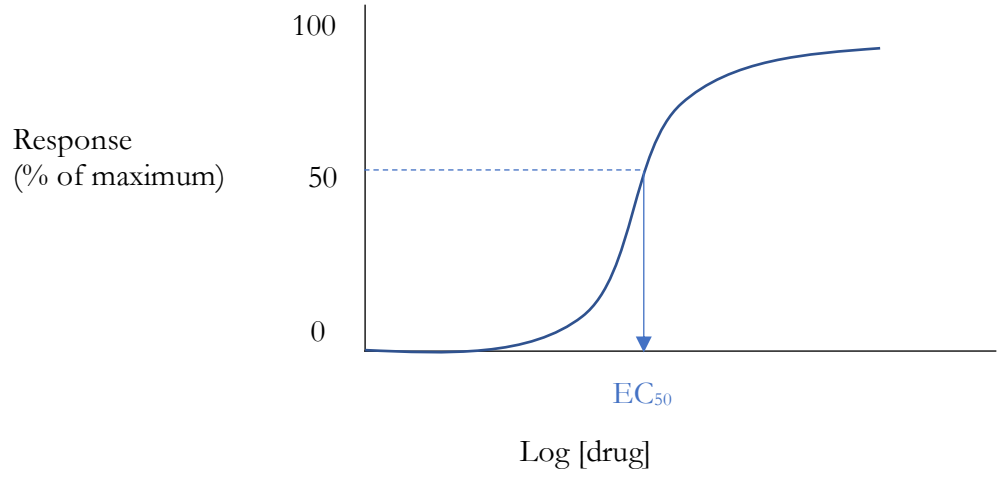

Figure 1.7 Concentration-effect curve of the intensity of the response against the logarithm of the drug concentration, and position of the EC_{50}, a common measurement of potency. Adapted from *(28)*.

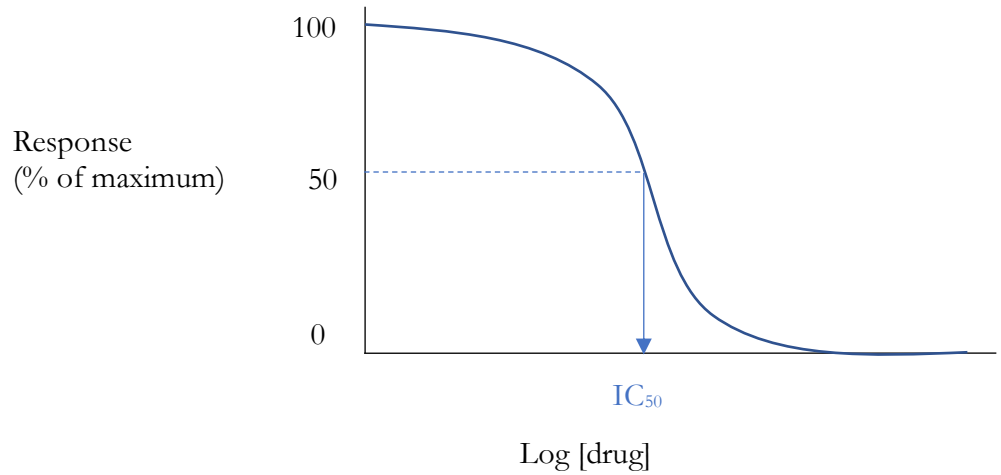

Figure 1.8 Concentration-effect curve of the intensity of the response against the logarithm of the drug concentration, and position of the IC_{50}, a common measurement of potency.

Potency is influenced by the affinity and the efficacy of the drug, and the receptor density *(28)*. Affinity refers to the binding of the drug to a receptor as a result of chemical forces, which can be quantified by the equilibrium dissociation constant, K_d, of the drug-receptor complex, and corresponds to the concentration achieving 50% receptor binding site occupancy *(28, 41)*. The binding affinity of competitive antagonists can be quantified with the inhibition constant, K_i *(36)*. Measurements of IC_{50} and EC_{50} depend on experimental conditions such as the concentration of agonist used in the assay, and therefore they do not directly indicate binding affinities *(22, 35)*.

However, determining affinities may be more time-consuming, and less indicative of tissue and organ-level effects than potency measurements *(42, 43)*.

Whilst it is clear that measurements such as K_i, EC_{50}, IC_{50} etc. are different from one another, in practice, different affinity and potency measures have been pooled in order to maximise the amount of usable bioactivity data derived from varied sources, such as for modelling, and one study found that this only moderately increases noise in the combined data *(18, 40, 44)*.

Polypharmacology, or compound promiscuity, refers to a single compound interacting with multiple protein targets, i.e. the extent of affinity for multiple targets *(45, 46)*. For example, clinical candidates and marketed drugs are generally active at more than one target; one study found that 35% of marketed drugs are active (< 10 μM) against 2 to 5 targets, and most other drugs against a higher number, although the extent of polypharmacology varies between drugs *(46, 47)*. While the unintended interaction of drugs with off-targets such as hERG can cause adverse drug effects, it is recognised that the polypharmacology of various approved drugs is essential to their therapeutic efficacy *(48, 49)*. For example, the action of antipsychotic drugs at multiple protein targets is thought to play a role in their efficacy *(50)*. Similarly, some oncology drugs achieve a beneficial synergistic effect by inhibiting multiple targets simultaneously *(48, 49, 51)*. Furthermore, it might be possible to reduce adverse effects mediated through on-target interactions when drugs act through multiple pathways simultaneously, as is the case for some analgesics *(48)*. In conclusion, while it remains important to avoid unintended off-target activities, polypharmacology is also of increasing interest for improving therapeutic efficacy and safety *(48, 49, 51)*.

1.4 Overview of safety assessment during drug discovery and development

The discovery and development of drugs is a lengthy process, of which some of the traditional stages are illustrated in Figure 1.9. In the earliest phases of drug discovery, thousands or millions of molecules may be screened for their activity against a biological target of interest e.g. in a high-throughput screening campaign. These are then narrowed down to a limited set of hit compounds to be further optimised into lead compounds, before a final candidate is chosen for the first animal studies and ultimately clinical studies involving human participants *(52–55)*. After marketing authorisation, potential safety issues are monitored in post-marketing surveillance efforts *(56)*.

Figure 1.9 Traditional phases of drug discovery and development
Adapted from *(22, 53)*.

The following sections will further introduce the safety-related studies in each of the drug discovery and development phases, focusing on *in vitro* pharmacological profiling and the analysis of post-marketing surveillance data, since these are of direct relevance to the current work.

1.4.1 Early drug discovery

1.4.1.1 In silico *and* in vitro *screening*

In the 1990s, several drugs were withdrawn from the market due to AEs, for example the appetite suppressant fenfluramine was associated with heart valve disease and the antihistamine astemizole carried an increased risk of arrhythmia *(57)*. Upon detailed analysis, these adverse effects were found to be related to specific interactions of the drugs with human targets: agonism at the cardiac serotonin receptor 2B (5-HT_{2B}) was found to induce heart valve disease and inhibition of hERG causes prolongation of the QT interval on the electrocardiogram, with the potential to cause life-threatening arrhythmias *(58, 59)*. These discoveries led to the introduction of additional regulatory requirements for both *in vitro* testing, such hERG channel inhibition, and *in vivo* testing, i.e. a battery of mandatory safety pharmacology investigations in animals *(15, 59)*. These developments, together with the high cost of late-stage failures, also motivate the wider trend to assess potential safety issues early in drug discovery during lead selection and optimisation *(15, 19, 52)*. *In silico* and *in vitro* techniques play an important role in this 'front-loading' *(15, 52)*.

In silico models exist to predict genotoxicity, protein target interactions such as with hERG and the bile salt export pump (BSEP), and sometimes more complex endpoints such as phospholipidosis that have some dependence on physicochemical parameters *(52)*. The types of studies done *in vitro* include cytotoxicity, genotoxicity, interaction with metabolic enzymes, as well as tissue-specific cell culture models *(52)*. However, one of the most widely used techniques is *in vitro* secondary pharmacological profiling, which refers to screening compounds against a variety of protein targets with possible links to adverse effects *(19, 20, 22)*. Some screening panels are dedicated to specific

target classes such as kinases or ion channels, or on establishing the general promiscuity, which has been associated with toxicity *(19, 60)*. *In vitro* secondary pharmacology screening has also been referred to as selectivity screening, pharmacological profiling, and *in vitro* safety profiling *(19, 20, 22)*.

1.4.1.2 Published secondary pharmacology screening panels

A number of safety pharmacology profiling panels have been published (Table 1.3). Generally, the targets included on such panels include G protein-coupled receptors (GPCRs), ion channels, enzymes, transporters and nuclear receptors. One panel published by Bowes et al. consists of 44 proteins which was compiled as the consensus set of targets screened by major pharmaceutical companies, intended to be used as a minimal set of targets to be screened *(19)*. This panel is also available commercially as the Cerep-44 panel *(61)*. Whitebread et al. compiled a set of 36 targets focused on cardiovascular toxicity *(20)*. The panel used at Abbvie includes over 70 targets *(62)*, whereas a list of targets used at Novartis has also been published *(63)*. Authors from Amgen suggested a panel of 70 targets for application in pharmaceutical development *(64)*. More recently, an optimised panel of 50 targets was published that was tailored to screening needs at Roche *(65)*. The specific targets in three published panels are shown by target class in Table 1.4. The overlap in targets between these three panels is roughly between 30 and 80%.

Table 1.3. Overview of published safety target panels

Publication	**Origin**	**Number of targets**	**Annotated biological effects, phenotypes or ADRs**	**Reference**
Bowes et al. (2012)	Minimal consensus panel from four companies / Cerep-44	44	Yes, with references	*(19)*
Lounkine et al. (2012)	Novartis	73	No	*(63)*
Whitebread et al. (2005)	Novartis – cardiovascular	36	Yes, without references	*(20)*
Lynch et al. (2017)	Abbvie – focus on pathology	70	Yes, with references	*(62)*

Deaton et al. (2019)	Amgen	70	Yes, partial references	*(64)*
Bendels et al. (2019)	Roche	50	No	*(65)*

Regarding the selection of targets included on safety target panels, the focus in most is on human targets although targets from other organisms are sometimes included *(20, 62)*. One cited important criterion for including targets is an established association with clinical ADRs, where serious and frequent ADRs are prioritised *(66)*. This results in a prominent representation of targets explored as therapeutic drug targets, such GPCRs (Table 1.4), since there is more clinical experience and knowledge of such targets *(19, 20, 66, 67)*. The emphasis on severe effects is reflected in the focus on targets related to the cardiovascular, respiratory, and nervous systems *(19, 64)*. In contrast to requiring established associations with clinical effects, Bendels et al. decided to focus on the general promiscuity across targets, including a large share of targets that lacked established knowledge of adverse effects *(65)*. However, Jenkinson et al. suggest that including targets without strong evidence of *in vivo* effects may have questionable value for decision-making *(22)*.

Another consideration for target inclusion is the redundancy in active compounds between targets, for instance, often a target family representative is included rather than multiple members of a family *(19, 64, 68)*. A cut-off of 80% ligand overlap or protein sequence similarities have been used for this *(64, 66)*. The optimised panel for screening at Roche was derived by statistical analysis of assay hit rates and iterative comparison of the overlap of active ligands between assays, in order to select the most non-redundant set of assays *(65)*.

Technical aspects of *in vitro* assay development also influence the selection of targets, for example ion channels are more difficult to screen in a high-throughput fashion, resulting in a historic underrepresentation of ion channels on safety pharmacology panels *(19)*.

Table 1.4 Targets in three published safety target panels by target class. The target class is the second level of the ChEMBL target hierarchy if available otherwise only the first level is shown.

Target class	**Bowes et al.** *(19)*	**Lynch et al.** *(62)*	**Whitebread et al.** *(20)*
Electrochemical transporter	• Dopamine transporter • Norepinephrine transporter • Serotonin transporter	• Dopamine transporter • Norepinephrine transporter • Serotonin transporter	None
Family A G protein-coupled receptor	• Adenosine A2a receptor • α-1a adrenergic receptor • α-2a adrenergic receptor • β-1 adrenergic receptor • β-2 adrenergic receptor • Cannabinoid CB1 receptor • Cannabinoid CB2 receptor • Cholecystokinin A receptor • δ opioid receptor • Dopamine D1 receptor • Dopamine D2 receptor • Endothelin receptor ET-A • Histamine H1 receptor • Histamine H2 receptor • κ opioid receptor	• Adenosine A1 receptor • Adenosine A2a receptor • Adenosine A2b receptor • Adenosine A3 receptor • α-1a adrenergic receptor • α-1b adrenergic receptor • α-2a adrenergic receptor • α-2b adrenergic receptor • Angiotensin II type 2 (AT-2) receptor • β-1 adrenergic receptor • β-2 adrenergic receptor • Bradykinin B2 receptor • Cannabinoid CB1 receptor • Cannabinoid CB2 receptor • Cholecystokinin A receptor	• Adenosine A1 receptor • Adenosine A2a receptor • Adenosine A3 receptor • α-1a adrenergic receptor • α-1b adrenergic receptor • α-2a adrenergic receptor • α-2b adrenergic receptor • α-2c adrenergic receptor • β-1 adrenergic receptor • β-2 adrenergic receptor • Bradykinin B1 receptor • Bradykinin B2 receptor • Dopamine D1 receptor • Endothelin receptor ET-A • Endothelin receptor ET-B

	• μ opioid receptor • Muscarinic acetylcholine receptor M1 • Muscarinic acetylcholine receptor M2 • Muscarinic acetylcholine receptor M3 • Serotonin 1a (5-HT1a) receptor • Serotonin 1b (5-HT1b) receptor • Serotonin 2a (5-HT2a) receptor • Serotonin 2b (5-HT2b) receptor • Vasopressin V1a receptor	• δ opioid receptor • Dopamine D1 receptor • Dopamine D2 receptor • Endothelin receptor ET-A • Endothelin receptor ET-B • Histamine H1 receptor • Histamine H2 receptor • κ opioid receptor • Melatonin receptor 1B • μ opioid receptor • Muscarinic acetylcholine receptor M1 • Muscarinic acetylcholine receptor M2 • Muscarinic acetylcholine receptor M3 • Muscarinic acetylcholine receptor M4 • Muscarinic acetylcholine receptor M5 • Neurokinin 1 receptor • Neurokinin 2 receptor • Neuropeptide Y receptor type 1 • Platelet activating factor receptor • Prostanoid EP2 receptor • Serotonin 1a (5-HT1a) receptor • Serotonin 1b (5-HT1b) receptor • Serotonin 2a (5-HT2a) receptor	• Ghrelin receptor • Histamine H3 receptor • Muscarinic acetylcholine receptor M1 • Muscarinic acetylcholine receptor M2 • Muscarinic acetylcholine receptor M3 • Muscarinic acetylcholine receptor M4 • Neuropeptide Y receptor type 1 • Serotonin 2b (5-HT2b) receptor • Serotonin 4 (5-HT4) receptor • Thromboxane A2 receptor • Type-1 angiotensin II receptor • Vasopressin V1a receptor • Vasopressin V1b receptor

		• Serotonin 2b (5-HT2b) receptor • Serotonin 2c (5-HT2c) receptor • Serotonin 4 (5-HT4) receptor • Serotonin 7 (5-HT7) receptor • Type-1 angiotensin II receptor • Urotensin II receptor • Vasopressin V1a receptor	
Family B G protein-coupled receptor	None	• Calcitonin gene-related peptide type 1 receptor • Vasoactive intestinal polypeptide receptor 1	• Calcitonin gene-related peptide type 1 receptor
Hydrolase	• Acetylcholinesterase	• Acetylcholinesterase • Sodium/potassium-transporting ATPase α-1 chain	None
Ligand-gated ion channel	• Acetylcholine receptor protein αchain • GABA receptor α-1 subunit • Glutamate (NMDA) receptor subunit zeta 1 • Neuronal acetylcholine receptor protein α-4 subunit • Serotonin 3a (5-HT3a) receptor	• GABA receptor α-1 subunit • Glutamate (NMDA) receptor subunit zeta 1 • P2X purinoceptor 1 • P2X purinoceptor 2 • P2X purinoceptor 3 • P2X purinoceptor 4 • P2X purinoceptor 5	• Acetylcholine receptor protein α chain

		• P2X purinoceptor 6 • P2X purinoceptor 7 • Serotonin 3a (5-HT3a) receptor	
Membrane receptor	None	• σ opioid receptor • Translocator protein	None
Nuclear receptor	• Androgen Receptor • Glucocorticoid receptor	• Androgen Receptor • Glucocorticoid receptor • Peroxisome proliferator-activated receptor gamma	None
Oxidoreductase	• Cyclooxygenase-1 • Cyclooxygenase-2 • Monoamine oxidase A	• Cyclooxygenase-2 • Monoamine oxidase A	None
Phosphodiesterase	• Phosphodiesterase 3A • Phosphodiesterase 4D	• Phosphodiesterase 3A	None
Primary active transporter	None	• Sodium/potassium-transporting ATPase α-1 chain	None
Protease	None	• Angiotensin-converting enzyme	None
Secreted protein	None	• TNF-α	None

Voltage-gated ion channel	• hERG • Sodium channel protein type V αsubunit • Voltage-gated L-type calcium channel α-1C subunit • Voltage-gated potassium channel subunit Kv7.1 (+ Voltage-gated potassium channel β subunit Mink)	• Potassium channel, inwardly rectifying, subfamily J, member 11 • Small conductance calcium-activated potassium channel protein 1 • Small conductance calcium-activated potassium channel protein 2 • Small conductance calcium-activated potassium channel protein 3 • Sodium channel protein type V α subunit • Voltage-gated L-type calcium channel α-1C subunit	• hERG • Potassium channel, inwardly rectifying, subfamily J, member 11 • Sodium channel protein type V α-subunit • Voltage-gated L-type calcium channel α-1C subunit

Screening compounds at a single fixed concentration, such as 10 μM, and observing the percentage inhibition is more cost-effective than determining the full dose-response curve to determine the IC_{50}, so this is often used in safety target screening *(19, 22, 62, 65)*. Often only target binding is established, rather than the agonist or antagonist mode, for instance using radioligand binding assays *(19, 20, 22, 62)*. Such binding assays primarily detect antagonists, and while in practice most drugs with binding affinity are antagonists, there are some exceptions such as 5-HT_{2B} agonism *(22, 62, 66)*. When using binding assays, target interaction and dose-response can subsequently be investigated in more detail *(22, 60, 62, 65)*. Other safety target panels do include agonist and antagonist modes, or consist of a combination of functional and binding assays *(60)*.

1.4.1.3 Remaining questions around target selection and interpretation

In early drug discovery, secondary pharmacology screening is widely used to flag potential adverse effects in large sets of compounds and to rank and (de)select compounds for further study *(19, 62)*. Later on, the results are used to design further preclinical and clinical studies *(19, 62, 69)*. However, despite their widespread use, there are remaining questions regarding a range of aspects of the current secondary pharmacology screening panels.

Firstly, the rationale for the selection of individuals targets is not consistently included on published panels, as illustrated by only two of the published panels including full references to supporting evidence on which target inclusion is based (Table 1.3). From the overlap between different panels it is also clear that companies have made different choices regarding targets to be screened (Table 1.4) *(22)*. For example, the panel by Lynch et al. includes one melatonin receptor, whereas Bowes et al. and Bendels et al. specifically did not include any melatonin receptors due to lack of evidence of adverse effects *(19, 65)*. Even in regulatory submissions, authors from the US FDA stated that "the panels of targets that are employed vary widely and are often selected without justification or a description of their relevance to human safety" *(70)*. In the panel reported by Bendels et al. *(65)*, most targets were included despite not having established evidence of safety implications while other targets were specifically excluded from the list of Bioprint targets on which the panel was based, due to lack of clinical relevance *(65)*. Other differences in target selection may be due to different experiences within organisations, therapeutic areas, and primary targets of compounds studied *(22)*. Thus, overall, there is no full agreement about the target set to use.

Secondly, there is a lack of information on how to interpret secondary pharmacology screening results and use the findings in subsequent decision-making. Comprehensive qualitative and quantitative annotations are considered important for interpretation and decision-making *(22, 62, 66)*. Off-target activities may give an early idea of potential susceptibilities to AEs e.g. in certain patient groups *(47)*, and therefore the results influence *in vivo* safety assessment and endpoints to monitor during clinical trials *(19, 66)*. For example, inhibition of phosphodiesterase 3 (PDE3) is linked to an increase in cardiac contractility that may have a negative impact in congestive heart failure *(47, 62)*. Thus, if this is relevant to the intended therapeutic area of the molecule or candidate, secondary screening results for this target can aid the compound selection *(47)*. However, only few of the published panels include full annotations of the *in vivo* biological effects, phenotypes, or associated ADRs (Table 1.3) *(22, 62, 66)*. Krejsa et al. reported that while secondary pharmacology screening results are used to design subsequent studies "the interpretation of off-target binding results is usually based on the personal experience of the development research team" *(69)*. Thus, additional, more widely available information of target-associated effects to support decision-making based on safety pharmacology would be valuable.

A third limitation of secondary pharmacology panels includes the lack of quantification of the correspondence between *in vitro* secondary pharmacology activities and *in vivo* effects. In early screening, the most important question is how likely an adverse effect is given an observed *in vitro* potency *(71)*. Therefore, Pollard et al. point to the Positive Predictive Value (PPV) and Negative Predictive Value (NPV) as the relevant measures for (de)selecting compounds; these measures identify the fraction true positives in terms of *in vivo* effects among all compounds active *in vitro*, and the fraction of true negatives among all inactive compounds, respectively *(71)*. Depending on the bioactivity threshold (0.3, 1, 3, 10, 30, 100, 300, and 1000 μM) for *in vitro* hERG inhibition, Pollard et al. report PPVs ranging from 1.00-0.29 and NPVs between 0.94 and 0.78 with respect to QT interval prolongation in humans *(71)*. To use this information for decision-making, Pollard et al. suggested that based on the low risk of *in vivo* effects for compounds with low-potency at hERG, compounds can be progressed at early stages *(71)*. Similarly, Krejsa et al. studied the usefulness of *in vitro* potencies at different safety targets for predicting clinical adverse effects, reporting the PPVs for different *in vitro* potency ranges *(69)*. For example, 80% of drugs with an IC_{50} under 0.1 μM at the muscarinic receptor M1 listed tachycardia on the drug label *(69)*. Similarly, the PPVs for the other IC_{50} bins were as follows (approximate values): 50% for 0.1-1 μM, 45% for 1-10 μM, 39% between 10-50 μM, and 25% for drugs with an IC_{50} > 50 μM *(69)*. Similar concentration-response relationship were reported for the relationships between histamine H1 and

somnolence, and between dopamine D1 and tremor or extrapyramidal symptoms *(69)*. Krejsa et al. suggested that analyses such as these can help to identify those targets with the strongest statistical links to ADRs, supporting the quantitative interpretation of *in vitro* pharmacological activities as well as the selection of targets for *in vitro* screening and computational model development *(69)*. However, beyond these two articles reporting on only few target-AE combinations, there is a lack of publicly available information on the predictive values of secondary pharmacology activities, with none of the published panels listed in Table 1.3 including any quantitative measures of association. Without quantitative information on the likelihood of *in vivo* effects, there is a risk that too many candidates are flagged if assays have high false positive rates, and potentially safe compounds are not taken forward, which would hamper the discovery process *(52, 66)*. A similar lack of information on the predictive value of other nonclinical safety assessments, and quantitative knowledge to inform decision-making, such as safety margins, and has also been identified *(19, 22, 24)*. Overall, there is a need for greater understanding of the translation of *in vitro* activities to *in vivo* effects and quantitative information on this relationship *(19, 66)*.

1.4.1.4 Additional uses of secondary pharmacology screening results

Apart from ranking and (de)selecting candidates for progression, one of the other main purposes of safety target screening is to inform structure-activity relationships that can be used in attempts to eliminate the off-target activities using medicinal chemistry *(19, 22, 66, 72)*. Therefore, structure-activity relationships may be built using the results *(22)*. The results of secondary pharmacology screens also feed into the development of *in silico* models for target activity that may become alternatives to *in vitro* profiling *(19, 60, 72)*.

In the lead optimisation phase, predictions for the therapeutic plasma concentration can be made, which allows teams to focus on achieving the widest safety margins *(19)*. A safety margin is the ratio of the AC_{50} in an *in vitro* assay and the (predicted) free plasma concentration *(19, 22)*. It is emphasized that activity on safety target panels does not necessarily mean that an ADR will be observed, since as described in the earlier, the ultimate effects of drugs also depend on the pharmacokinetics *(20, 66)*. Thus, integration with the free plasma concentration is considered useful and important for the interpretation of off-target activities and the translation to adverse effects in humans *(19, 20, 22, 57, 66, 69, 73)*. It is also recommended to focus on this kind of quantitative, integrated risk assessment in regulatory submissions *(19, 70)*.

Another use of secondary pharmacology screening is to rationalise species differences observed in preclinical toxicity studies or discrepancies between the effects observed in animal studies and clinical trials *(22, 66)*. This also highlights one of the potential advantages of *in vitro* pharmacological profiling, which is that it focuses on human targets, since some adverse effects may be difficult to examine in animals, e.g. suicidal intent or hallucinations *(15, 67, 74)*. Muller et al. identified 10 protein targets associated with suicidal intent and behaviour, showing that *in vitro* target activities can be associated with complex AEs *(75)*.

1.4.1.5 Compound promiscuity

Compound promiscuity has been found to correlate with toxicity, e.g. studies found that promiscuity is correlated with *in vitro* cytotoxicity *(76)* and *in vivo* toxicity findings in animal studies *(76–78)*. Similarly, another study found that candidates that were terminated are more often promiscuous than marketed drugs *(79)*. The knowledge that compound promiscuity is linked to several physicochemical properties such as lipophilicity and basicity is being used in drug discovery to reduce undesired promiscuity *(19, 45, 76, 78)*. Another trend is the use of small *in vitro* pharmacology panels in early drug discovery, similar to the safety target panels introduced earlier, but these promiscuity panels are run separately and include a much smaller set of targets. For example, Sameshima et al. described a panel of 8 targets from diverse target families which correlated well with promiscuity measured in larger panels *(76)*. Such small-scale promiscuity panels can be used early on to filter out promiscuous compounds *(76)* and is considered separately from multi-target activity in larger *in vitro* pharmacology screens *(78)*. The former establishes promiscuity for a compound while not having knowledge of the precise targets and potential consequences, whereas the latter focuses on identifying secondary targets and their potential consequences for specific ADRs *(78)*.

1.4.2 Preclinical phase

In the preclinical phase, the pharmacokinetics and toxicity of drug candidates are investigated in animals, such as during repeat-dose studies over several weeks or months *(55)*. A range of doses is included and generally the number of animals per dose group is usually 20 or fewer *(55)*. Aspects investigated include body weight, electrocardiography, haematology, clinical chemistry, histopathology of the organs, as well as carcinogenicity and reproductive toxicology studies *(55)*. The doses used are generally higher than the intended therapeutic dose *(55)*. Separately, safety pharmacology studies investigate pharmacologically-mediated effects on important physiological functions - cardiovascular, respiratory and central nervous systems - often after a single dose in

the therapeutic range or above *(15, 55)*. Both toxicology and safety pharmacology studies follow strict sets of guidelines and regulatory requirements with respect to study design and endpoints assessed *(55)*.

A number of studies performed statistical analyses of large datasets to study the concordance between preclinical and clinical AEs. Toxicities in animals predict a reasonable fraction of clinical adverse effects, but they remain imperfect, especially in their negative predictive value, that is the absence of animal toxicity predicting the absence of ADRs in humans *(74, 80, 81)*. The predictive value of animal studies also varies by organ system *(74, 80, 81)*.

1.4.3 Clinical phase

During clinical trials, drug candidates are tested in humans for therapeutic efficacy and safety. The process usually takes place according to the following four phases *(54)*. In phase I clinical trials, the drug is tested in around 25-100 healthy volunteers for tolerability, pharmacokinetics, and safety *(54)*. The efficacy and safety are investigated in phase II trials and phase III trials, including hundreds to thousands of subjects with the relevant condition *(54, 55)*. Phase IV refers to post-marketing surveillance, often based on over 10,000 subjects with the condition in the population (see section 1.4.4 below) *(54, 55)*. Clinical trial phases I-III usually include one or more control groups who are given a placebo or an active comparator drug *(54, 82)*. Blinded, controlled trials help to control bias and confounding factors and are the gold standard for evaluating new drugs *(54, 82, 82)*.

The safety aspects studied during clinical trials include AEs, physical examinations, laboratory tests, vital signs, electrocardiography, and hospitalisations *(82)*. The extent of safety monitoring generally decreases with these phases, for example phase I subjects are monitored in the clinic but phase III patients would only have regular visits to their clinician *(55)*. Significant differences in the incidence and severity of AEs between the control and treatment groups are identified *(54, 82)*. AEs observed during clinical trials are commonly recorded in the MedDRA vocabulary *(56, 82)*.

Clinical trials have limitations in terms of their duration, diversity of patients included, and sample size *(14, 83)*. If the duration of clinical trials does not reflect the chronic use of the medication, clinical trial data can underestimate the true level of adverse drug effects *(82)*. The diversity of patients included is also limited, for example elderly patients and those with co-morbidities are

typically excluded or underrepresented in clinical trials *(54, 82)*. Moreover, some AEs are rare, e.g. occur in less than 1 in 1000 patients, and the sample size of clinical trials may not detect such events *(3, 82)*.

In addition to inherent limitations to clinical trials, there are also concerns about their reporting. Although it is required by US law that clinical trial results, including serious AEs, are reported on e.g. ClinicalTrials.gov, a recent study found that results for 35% of registered trials have not been reported *(84)*, and even when results are reported, inconsistencies in the serious AEs reported for the same study in peer-reviewed literature versus on ClinicalTrials.gov were found to occur in 35% of cases, in one survey *(85)*. Once drugs are marketed, information on adverse effects is included in the patient information leaflet, as well as on the prescribing information for health professionals *(56, 86)*. This latter information has been subject to text-mining to create datasets for research *(87)*, as will be discussed in more detail later.

1.4.4 Post-marketing phase

Monitoring of AEs continues beyond marketing authorisation *(3)*. Most commonly, this occurs in the form of spontaneous reporting, i.e. healthcare providers and consumers can voluntarily report case reports of AEs and suspected ADRs to regulatory authorities or directly to pharmaceutical companies *(56, 88–90)*. Data from post-marketing surveillance are of interest because they reflect the more realistic and diverse real-world population, and are based on the exposure of a greater number of subjects. The individual reports and databases containing collections of reports are then analysed for signals *(3, 88)*.

A signal has been defined by the World Health Organisation (WHO) as "reported information on a possible causal relationship between an AE and a drug, the relationship being unknown or incompletely documented previously" *(3)*. Both detailed clinical reviews of individual submitted reports and large-scale statistical data mining are used to evaluate potential signals *(88, 89)*. Health authorities can act upon safety concerns with regulatory actions such as requiring changes the drug label with respect to warnings or contraindications, implementing restrictions on drug use, informing the public of possible risks, or withdrawing the drug from the market *(3, 56, 89)*.

One of the largest spontaneous reporting databases is the US FDA Adverse Event Reporting System (FAERS) database, and other large databases are VigiBase, maintained by the WHO, and EudraVigilance, maintained by the European Medicines Agency (EMA) *(91–93)*. Pharmaceutical

companies also maintain their own databases of AEs *(88)*. Most pharmacovigilance organisations use disproportionality analyses for signal detection *(88, 94)*. These types of analyses determine whether the rate of the AE reporting with a drug of interest is higher than would be expected based on the background rate of AE reporting across all other drugs in the database *(94)*.

1.4.4.1 Biases in spontaneous reporting data

Since spontaneous reporting data are based on voluntary reporting, the sampling structure is uncertain; the numbers of patients who have taken the medicine cannot be derived from spontaneous reporting data *(3, 90)*. Therefore, spontaneous reporting systems cannot be used to estimate incidence rates of AEs, and the term 'reporting rates' are used instead *(90)*.

In randomised controlled studies, randomisation should reduce systematic differences between the treatment groups, but this is not the case for drug prescriptions in the general population from which spontaneous reporting data is derived *(95, 96)*. In routine clinical care, drug prescription is based on the patient's underlying conditions, resulting in differences between patients prescribed certain drugs and other patients *(97)*. As a result, spontaneous reporting data suffers from potential biases *(90, 96)*. One source of bias is sampling variance, which is the variation in reporting rates of AEs across e.g. time, drugs, and the type of AE *(98, 99)*. Examples of such variation are the greater reporting of AEs considered more important or serious by health care professionals and increased reporting for drugs covered in news reports *(9, 73)*. Another type of bias, reporting bias, results from the underreporting of AEs, which is thought to be widespread, with one meta-analysis finding rates of underreporting between 82-94% *(9, 100)*.

Another type of bias, sampling bias, can result from the existence of confounding factors. A confounding factor is a covariate such as a co-medication, co-morbidity or indication that is associated with exposure to the drug of interest *(101, 102)*. For example, if two drugs are frequently co-prescribed, but only one of them causes a particular side effect, statistically both drugs will be associated with the side effect *(101)*.

One of the main confounding factors in FAERS is drug indication, since treatment choices in the population are based on the conditions affecting the patients, and this can result in indication bias *(97)*. In such cases, it can be difficult to determine whether observed AEs are related to exposure to the drug, or underlying factors such as the disease or age *(97)*. In studies of FAERS it has been observed that drugs tend to be statistically associated with AEs related to the disease that the drugs

intend to treat. For example, it happens that drugs against type-2 diabetes are associated with high blood glucose *(73, 98)*. However, rather than this being a consequence of drug use, it is likely a consequence of the disease *(98)*. One study estimated that up to 5% of all reports in FAERS list the drug indication as an AE *(73)*. Perhaps the report submitters confuse the disease indication and the AE *(73)*, however, indication-related AEs are frequently due to the underlying disease and its symptoms. For example, a case studies of different drugs used for hypertension showed that around double the rate of cardiac and vascular disorders were reported in post-marketing data compared to clinical trials, which can be attributed to the disease being treated rather than the drug *(47)*.

Statistical methods have been developed to reduce potential biases in observational data such as post-marketing databases, and a further section below (1.7.2) is devoted to these in more detail.

1.5 The study of target-adverse event relationships using data analysis

The introduction so far has outlined the drug discovery process and the various types of studies performed at each stage. Advancing through each stage in drug development, more information of the *in vivo* drug effects becomes available. Knowledge of target-AE relationships is useful at different stages, but especially in early drug screening and lead optimisation, when safety target panels can be screened and computational models for off-targets can be used *(19, 66)*.

The analysis of marketed drugs has been important in the discovery and annotation of currently known safety-related targets, as illustrated by the previously mentioned examples of 5-HT_{2B} and hERG *(66, 67)*. However, there are remaining questions around secondary pharmacology screening and there is a need for further studies to address the lack of quantitative annotations and knowledge of the predictive values of targets on screening panels. One way to study target-AE relationships is through the analysis of empirical datasets, such as measured *in vitro* bioactivities and AEs observed in humans, and this is the focus of the current work.

The main types of data needed for such analyses are links between drugs and their pharmacological activity at protein targets - drug-target links – and links between drugs and AEs – drug-AE links. This is illustrated conceptually in Figure 1.10. The next sections will introduce sources of such data and applicable methods, and discuss previous literature of target-AE associations.

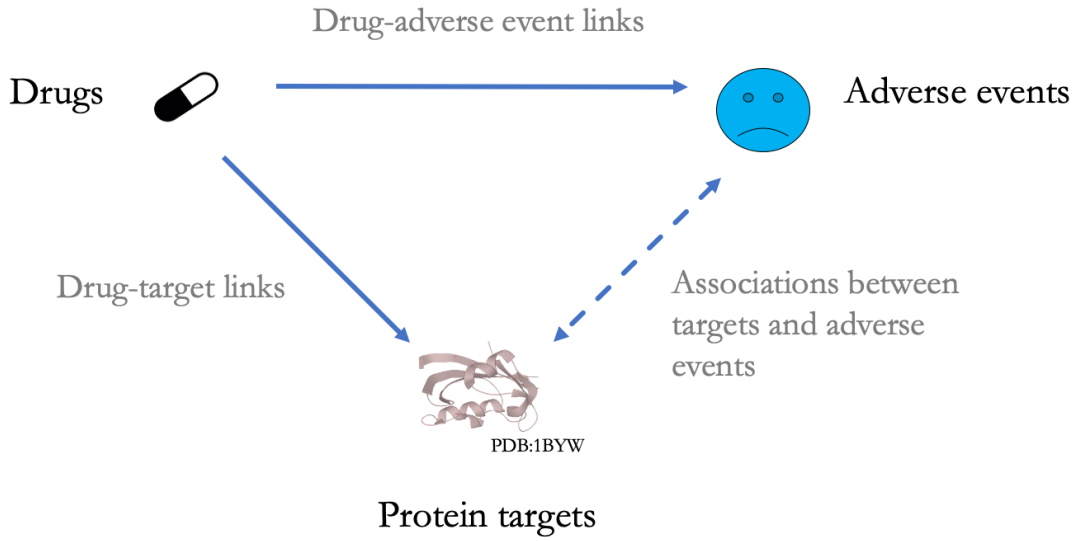

Figure 1.10 Associations between targets and adverse drug effects can be derived based on datasets of drug-AE links and drug-target links.

In addition to data analysis of *in vitro* pharmacological activities, other methods to discover targets that mediate adverse effects include knock-out studies and studies of phenotypes associated with genetic variation in humans *(64, 67, 103)*. These methods have also been used to inform the selection of proteins for safety target panels *(64, 67)*. For example, a recent study by Deaton et al. investigated correspondence between pharmacological modulation of targets and human genetic information of the target, and used this to inform targets to be included on a panel of 70 targets intended for pharmaceutical development *(64)*. The study revealed that the highest correspondence between pharmacological effects and genetic phenotypes occurs for AEs in the categories blood platelet disorders, seizures and coronary artery disease *(64)*. Another recent study by Nguyen et al. found a correlation between the effects of genetic variation in proteins that are drug targets and ADRs observed in clinical trials of drugs targeting these same proteins *(103)*. However, the discussion of data sources and previous literature below will focus on pharmacological and bioactivity data, given these are directly relevant to the current work.

1.6 Data sources for studying relationships between drugs, bioactivities, and adverse events

1.6.1 Large-scale measured and predicted bioactivities

Several databases provide large-scale measured bioactivity data. Among the largest publicly available ones is the ChEMBL database *(104)*. It provides *in vitro* bioactivity data that was manually

extracted from medicinal chemistry scientific literature, with additional curation of compound structures and information on approved drugs *(105)*. The ChEMBL database version 24 published in 2018 contains over 1.8 million unique compounds with measured bioactivity against 3,569 unique human proteins *(106)*. The activity types include IC_{50}, K_d, K_i, and percentage inhibition *(105)*. Except for a few subsets of the data, the majority of data is not a full matrix and is sparsely populated *(107)*. Negative data is underrepresented since this is often not included in scientific publications, thus ChEMBL contains a relatively high proportion of active compounds *(105, 106)*. ChEMBL also incorporates a selection of data from PubChem (making up ~40% of ChEMBL data), another large publicly available bioactivity database *(104, 105)*. PubChem focuses on storing results from high-throughput screening initiatives, and thus contains more full-matrix and negative data *(104, 105)*. Only the datapoints derived from dose-response assays in PubChem are incorporated in ChEMBL *(105)*. The ExCAPE-DB integrates datapoints from ChEMBL and PubChem and standardised the compound structures in a common way for use in chemogenomics studies *(108)*. Similarly, the Search Tool for Interacting Chemicals (STITCH) also collates a range data sources, including databases such as ChEMBL and DrugBank and datasets curated from scientific literature, into one integrated dataset *(109)*. DrugBank is a manually curated dataset of drug-target interaction data, as well as other data about approved drugs and clinical candidates *(110)*. However, numeric affinities and potencies are not readily available from DrugBank, since drug targets are provided as a list.

A relevant commercial source of bioactivity data is the Bioprint database developed by Cerep (now Eurofins) *(69)*. The database is a nearly full matrix of measured *in vitro* bioactivities of drugs and drug-like compounds against targets specifically related to safety *(45)*. In 2012, the database contained 2,413 compounds and 141 targets *(45)*.

Another effort to generate publicly available *in vitro* bioactivity data is the ToxCast and Tox21 initiative resulting from a collaboration between different US federal agencies *(111)*. The programme screened over 10,000 compounds, including drugs, in a range of assays focused on nuclear receptor interaction and higher-level cellular responses such as activation of stress-response pathways *(111)*.

It is acknowledged that public availability of measured bioactivity data, even for marketed drugs, is not comprehensive, with many missing datapoints *(63, 107)*. The availability of the above-described databases, amongst others, has facilitated the development of *in silico* ligand-based target

prediction models *(22, 107, 112)*. Such models generally use numeric descriptors of the compounds and machine learning models to predict the potency of a query compound or a probability of activity based on classification models *(112)*. The numeric descriptors are computed from the compound's two-dimensional structure and predictions are based on underlying similarity between the compound of interest and known ligands for the target of interest *(63, 113)*.

The target prediction tool 'Prediction IncluDinG Inactivity' (PIDGIN), which is used in this work, uses Random Forest models to predict the probability of activity based on RDKit Morgan fingerprints *(114)*. Morgan fingerprints are type of circular fingerprint that iteratively consider the atom environments in a given radius around atoms in a molecule *(115)* (Figure 1.11). Chemical substructures are condensed by a hashing function into features, the presence or absence of which can then be represented as a bit vector of a defined length describing a molecule *(115)*.

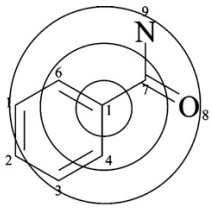

Figure 1.11 Circular fingerprints consider atom environments around a central atom. Reprinted with permission from D. Rogers, M. Hahn, Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **50**, 742-754 (2010) Copyright 2010 American Chemical Association.

PIDGINv3 is a collection of classification models that are trained on a dataset of known active molecules and a combination of measured inactive and putative inactive molecules from ChEMBL and PubChem *(116)*. A classification threshold, e.g. 10 μM, is typically used to define active and inactive molecules for each protein *(114)*. PIDGINv3 supports multiple classification thresholds per protein, namely 0.1, 1, 10 and 100 μM, with a separate model for each threshold *(116)*. Morgan fingerprints of 2,048 bits are calculated for the active and inactive molecules using a radius of 2 *(114)*. The Random Forest algorithm is used to train models that aim to relate chemical features from the fingerprints to compounds' biological activities *(117)*. The applicability domain of the PIDGINv3 models is defined using the reliability-density neighbourhood method, which considers the reliability and abundance of datapoints similar to the query compound in the training set *(118)*.

Other target prediction methods use different molecular fingerprints or machine learning methods. For example, the Similarity Ensemble Approach (SEA) uses the Daylight chemical fingerprints computed from the two-dimensional compound structure, to compare query compounds and known ligands of a target *(113)*. The method then uses a chemical similarity model to calculate the significance over similarities expected at random *(63, 113)*. Another target prediction workflow by Lampa et al. uses the signature molecular descriptors in a Support Vector Machine *(112)*. Similar to PIDGIN, Bosc et al. used the RDKit Morgan fingerprints in combination with Random Forest models and conformal prediction *(107)*.

1.6.2 Side Effect Resource (SIDER)

The Side Effect Resource (SIDER) is primarily derived from ADRs observed in clinical trials. The dataset contains a list of ADRs for 1,556 marketed drugs. These have been extracted via text-mining from the drug prescribing information, which are the documents prepared by drug manufacturers listing information about the approved drug for use by professionals *(18, 87, 119)*. The most abundant information in SIDER is the listing of adverse effects in combination with a drug without further quantification. For around 40% of datapoints, the frequency of the ADR in patients is available as a percentage of patients affected or description of the frequency. Only for 20% of datapoints with a frequency, the corresponding frequency in the placebo group is available for comparison *(87, 119)*. A few examples of datapoints from SIDER are shown in Table 1.5. Although SIDER also provides MedDRA LLTs in most instances, the PT is recommended for analysis *(47, 120)*.

Table 1.5 Example datapoints from SIDER version 4 for levocarnitine (STITCH identifier CID000010917) and bupropion (CID000000444), showing the listing of MedDRA PTs for drugs, different percentages of patients affected in different studies of the same drug (levocarnitine), and occasionally a frequency for the placebo group. In most cases, frequency information is not available, such as in the case levocarnitine-gastritis.

Data from the SIDER version 4 download files *(120)*.

Drug	**Adverse event (MedDRA PT)**	**Frequency (%) or description of frequency**
LEVOCARNITINE	Anaemia	12
LEVOCARNITINE	Anaemia	3

LEVOCARNITINE	Anaemia	5
LEVOCARNITINE	Anaemia	6
LEVOCARNITINE	Anaemia	Placebo: 3
LEVOCARNITINE	Gastritis	Not available
LEVOCARNITINE	Convulsion	Not available
BUPROPION	Amnesia	Rare
BUPROPION	Oedema	Frequent

1.6.3 FDA Adverse Event Reporting System (FAERS)

Of the different post-marketing databases maintained by health authorities and other organisations, FAERS is the only one publicly available as a complete database download, making it suitable for large-scale informatics research *(91–93)*.

FAERS is a collection of individual case reports submitted to the FDA by health care professionals, consumers, and drug product manufacturers *(89)*. Overall, around half of all reports are submitted by health care professionals, a further 40% by consumers, 3% by lawyers, and 9% are unspecified *(73)*.

An overview of the tables included in the FAERS database and the main types of information they contain is shown in Figure 1.12. The tables are linked by case and report identifiers.

DEMO (demographic)

- Age
- Sex
- Weight
- Reporter occupation, one of:
  - Physician
  - Pharmacist
  - Health-professional
  - Lawyer
  - Consumer
- Reporter country

DRUG (drugs)

- Drug name
- Dose
- Route
- Etc.

INDI (indications)

- Indications, i.e. diagnoses (MedDRA-coded) for reported drugs

THER (therapies)

- Drug start date
- Drug end date

REAC (reactions)

- Reaction (MedDRA-coded)
- Dates of reaction start and end

OUT (outcome)

- Patient outcome, one of:
  - Life-Threatening
  - Hospitalization – Initial or prolonged
  - Disability
  - Congenital Anomaly
  - Required Intervention to Prevent Permanent Impairment/Damage
  - Other Serious

RPSR (report sources)

- Source of report, one of:
  - Foreign
  - Study
  - Literature
  - Consumer
  - Health professional
  - User Facility
  - Company Representative
  - Distributor
  - Other

Figure 1.12 Overview of the main tables and content in FAERS, adapted from the documentation provided with the FDA quarterly release files from 2019 and related documentation
References: *(89, 91)*.

A typical report includes patient characteristics (demographics table), the drugs the patient was prescribed (drugs table), the indications/diagnoses for which the drugs were prescribed (indications table), AEs/reactions experienced (reactions table) and a categorical outcome (outcomes table). However, the level of completeness of the various fields in the database varies, e.g. not all reports list the patient demographics or dose *(17)*. For example, in a recent quarterly release of FAERS (faers_ascii_2019Q3), less than half of the reports include the patient age and weight *(91)*.

Drugs may also be classified as primary suspect, secondary suspect, concomitant, or interacting *(121)*. An example of the information of one report is shown in Table 1.6, listing the drugs and AEs, and one drug annotated as primary suspect for the AEs.

Table 1.6 The drugs used, indications, and AEs experienced listed on one case report in FAERS. Data extracted from FAERS Adverse Event Open Learning through Universal Standardisation (AEOLUS) *(121)*

Case ID (isr)	Drugs	Drug indications	Adverse events (MedDRA)
8310066	Loperamide; Leflunomide (primary suspect); Acetaminophen; Felodipine; Prednisolone; Lenalidomide; Zopiclone; Propiomazine	RHEUMATOID ARTHRITIS	BLOOD ALKALINE PHOSPHATASE INCREASED; BLOOD BILIRUBIN INCREASED; TRANSAMINASES INCREASED

Sometimes it happens that several versions of the same report exist in the database, for example because an updated report on the same case is submitted *(121, 122)*. Deduplication of such reports is not performed by the FDA, since it publishes the raw data without processing or analysis applied *(91)*. Although AEs and drug indications in FAERS are recorded in the MedDRA vocabulary, the drugs names are not standardised or mapped to external sources. The need for extensive pre-processing of FAERS has resulted in the publication of curated versions of FAERS which have duplicate case reports removed and drug names standardised *(121, 123, 124)*. This is important because such aspects can influence statistical analyses performed on FAERS *(73)*.

1.7 Statistical methods

1.7.1 Signal detection in spontaneous reporting data

Disproportionality methods are applied on databases such as FAERS to identify overrepresented combinations of drugs and AEs which may be signals *(97)*. Therefore, observed-to-expected ratios are determined for drug-AE combinations based on a contingency table as shown for an example in Table 1.7.

Table 1.7 Example of the contingency table used for calculating the disproportionality of the reporting of the AE 'injection site pain' with the use of etanercept. The background used is all other reports, i.e. reporting on drugs other than etanercept, in the database.
Example from Banda et al. *(121)*.

	Reports with *Injection site pain*	Reports without *Injection site pain*	Total
Reports with *etanercept*	30,793 (a)	647,134 (b)	$a+b$
All other reports	140,853 (c)	61,815,244 (d)	$c+d$

Different measures of disproportionality include the Proportional Reporting Ratio (PRR), Reporting Odds Ratio (ROR), Information Component (IC) and Empirical Bayes Geometric Mean (EBGM) *(88)*. For example, the PRR for injection site pain with etanercept according to the numbers in Table 1.7 would be calculated as follows *(125)*.

$$Proportional\ Reporting\ Ratio\ (PRR) = \frac{\frac{a}{(a+b)}}{\frac{c}{(c+d)}} = 19.98 \qquad (1.7.1)$$

The PRR compares the rate of AE reporting among the reports listing a drug of interest, corresponding to $a/(a+b)$, with the rate of AE reporting in background reports, which is $c/(c+d)$ *(97)*. Similarly, the ROR is also based on the concept of disproportionality but is a ratio of odds: the odds of the number of reports of a given AE with the drug of interest to the number of reports for the same AE with other drugs, i.e. a/c, over the odds of the number of reports with other AEs with the drug of interest to those without the drug of interest, i.e. b/d *(126)*. Therefore, the ROR is calculated as follows *(125, 126)*:

$$Reporting\ Odds\ Ratio\ (ROR) = \frac{(a/c)}{(b/d)} \qquad (1.7.2)$$

The IC is derived using a Bayesian framework by taking the logarithm (base 2) of the observed-to-expected ratio, resulting in a value similar to the $log_2 PRR$ *(97, 99, 125, 127)*:

$$Information\ Component\ (IC) = log_2\left(\frac{a}{\frac{(a+b)(a+c)}{(a+b+d+d)}}\right)$$

$$= log_2 \frac{a(a+b+c+d)}{(a+b)(a+c)} \qquad (1.7.3)$$

This incorporates the expected count E that is based on assuming no association between the drug and the AE *(99, 125)*:

$$E = \frac{(a+b)(a+c)}{(a+b+c+d)} \tag{1.7.4}$$

Hence, the IC represents the observed-to-expected ratio. The expected count is based on the number of reports of a specific drug in the database and the number of reports of a specific ADR in the database. In a Bayesian framework, the expected count represents the prior estimate of the probabilities *(127)*. This is compared to the probability of the ADR being listed on a report given a drug is listed on a case report, i.e. the posterior probability, which is based on the number of times a specific drug-ADR combination occurs in the database *(127)*. The basis of one method, the Bayesian Confidence Propagation Neural Network (BCPNN) is that as new reports are added to the database, previously calculated posterior probabilities are used as prior estimates, resulting in constantly updated estimates of the IC *(127)*. In addition, a confidence interval is calculated for the IC which helps to define the signals *(127)*. The EBGM is a similar measure of disproportionality also derived by Bayesian inference, and while it is also based on the observed-to-expected ratio, it adjusts the ratio towards 1, which is the null value, i.e. no association between the drug and the AE *(97)*. The advantage of this 'shrinkage', which is also already incorporated in the IC *(128)*, is that the resulting values are less extreme when the number of reports are very small *(97, 129)*. These adjustments prevent large numbers of positive signals and thus give more weight to associations with greater support in terms of the number of reports *(97, 129)*. In contrast, the PRR and ROR are less reliable when the number of reports is small *(97)*. Another approach to avoid false positive associations is to require a minimum number of reports for the drug-AE combination, such as at least three or five reports *(88)*. Examples of signal detection thresholds for disproportionality measures used by pharmacovigilance organisations are shown in Table 1.8, some of which incorporate a minimum number of reports.

Table 1.8 Examples of signal detection methods used by various organisations. *n*=number of reports.

Adapted from Candore et al. *(88)*.

Method	**Thresholds**	**Organisations using threshold**
PRR	PRR ≥ 2 and $\chi^2 \geq 4$ and $n \geq 3$	Bayer; AstraZeneca until 2009

	PRR ≥ 3 and $\chi^2 \geq 4$ and $n \geq 3$	Historic use at Medicines and Healthcare products Regulatory Agency (MHRA)
ROR	ROR lower bound 95% CI > 2 and $n \geq 5$	Medicines Evaluation Board
IC	IC lower bound of the 95% CI > 0	Uppsala Monitoring Centre
EBGM	EB05 ≥ 1.8 and $n \geq 3$ and EBGM ≥ 2.5	MHRA

Some measures of disproportionality have advantages in certain situations, such as the ROR performing best for early detection of AEs *(130)*. However, all these methods converge to the same result given enough drug-AE reports, e.g. most methods give similar results if the number of reports is greater than 3 *(97)*, and it has been suggested that the choice of algorithm can be based on the convenience of implementation *(88, 94)*.

Pharmacovigilance organisation have specialised data management systems for storing and running signal detection algorithms, and this software is not available to most academic researchers *(131)*. To improve the situation for researchers, Banda et al. *(121)* published a standardised version of FAERS named FAERS AEOLUS. In addition to standardised and mapped drug names and case report deduplication, FAERS AEOLUS includes precalculated PRRs and RORs *(121)*. While this provides an easily accessible source of drug-AE associations, potential biases have not been addressed in this dataset, which will be discussed in the next section.

1.7.2 Addressing potential biases in post-marketing reporting data

As introduced in section 1.4.4.1, observational data suffers from a range of biases due to sampling variance and confounding factors.

With respect to sampling variance, the methods described above incorporating shrinkage already address the issue of possible spurious associations arising when there are only few reports for a given drug or AEs, which makes reporting ratios more sensitive to sampling variance *(129, 132, 133)*. To counteract varying reporting rates due to media attention, Maciejewski et al. calculated

reporting rates in FAERS on a monthly basis and correlated this with reports in the media *(73)*. Observing whether the monthly reporting rates are still elevated outside of periods of media attention can help distinguish between true signals and signals due to stimulated reporting *(73)*.

With respect to confounding factors, disproportionality methods by themselves do not control for confounding factors *(94)*, but these factors can lead to the masking of true associations or the detection of false associations when using disproportionality methods *(132)*. Methods to reduce confounding factors include stratification, subgroup analysis, regression, and propensity scores.

1.7.2.1 Stratification and subgroup analysis

Stratification is a procedure in which the data is divided into groups, or strata, each with the same or a restricted range of the confounding factor *(134)*. The expected counts are then calculated for each of the strata and pooled to give an overall weighed expected count that is compared to the observed rate according to the Mantel-Haenszel procedure *(99, 134)*:

$$Adjusted\ observed\text{–}to\text{–}expected\ ratio = \frac{(a_1 + a_2)}{\left(\frac{(a_1 + b_1)(a_1 + c_1)}{(a_1 + b_1 + c_1 + d_1)} + \frac{(a_2 + b_2)(a_2 + c_2)}{(a_2 + b_2 + c_2 + d_2)}\right)} \quad (1.7.5)$$

where the values $a, b, c,$ and d correspond to those in Table 1.7, and the subscripts 1 and 2 might be male and female, for example, or another variable to be stratified on *(99)*.

In the related subgroup analysis, the disproportionality statistic is calculated for each of the strata separately and it is considered a signal if the statistic of any of the strata is significant *(134)*. Both stratification and subgroup analysis are effective in reducing the effects of confounding factors but subgroup analysis performed better for large databases when evaluated against a reference of known ADRs *(134)*. A limitation of both methods is that only a limited number of co-variates can be taken into account and it would not be feasible to include large numbers of co-medications at the same time, which occurs in FAERS *(94, 128)*.

1.7.2.2 Regression

Regression models can take into account multiple variables at the same time and can be used to adjust for confounding factors. For example, logistic regression can be used to model the odds of an ADR y being listed on a given AE report as the binary dependent variable:

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \quad (1.7.6)$$

where p is the probability of a report listing the ADR y, $\Pr(y = 1)$, and the explanatory variables x_1 to x_n are different other drugs listed on the report or other potential confounding factors, such as drug indications *(128, 135)*. The inclusion of confounding factors allows for their adjustment. For example, one study used logistic regression to investigate contrast agent-induced nephropathy, and considered the co-prescription of 200 nephrotoxic drugs in the regression, as well as including the patient age, patient sex and the year of the report *(135)*. The regression coefficients were converted to estimated Odds Ratios to identify which specific contrast agents were associated with nephropathy *(135)*. In another study, logistic regression models were used for all possible reported AEs, separately, as a means of signal detection in the database *(128)*. Therefore, identifying drugs with positive β coefficients for given ADRs enabled the listing of drug-AE combinations as signals, while confounding due to co-prescriptions was controlled *(128)*. Regression models can unmask associations that would remain undiscovered with other methods and prevent false associations that may be caused by confounding factors *(128, 135)*.

1.7.2.3 Propensity scores

Propensity scores are another method used to reduce biases in observational studies *(136)* and are used in this thesis. A propensity score is defined as "the conditional probability that a person will be in a treatment group, given his or her specific characteristics" *(136)*. True propensity scores are unknown in observational research, but may be estimated using available data *(137)*. The most common way to do so is by modelling the treatment status Z of a patient, where the patient is either treated (Z=1) or not treated (Z=0) with a particular drug, as dependent on the baseline patient characteristics x_1 to x_n in logistic regression *(137, 138)*:

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \quad (1.7.7)$$

where p is the probability of a patient being prescribed a drug, $\Pr(Z = 1)$, based on their baseline characteristics *(137)*. In contrast to the previous logistic regression (1.7.6), in which the ADR being reported is the dependent variable, now the dependent variable represents an intermediate step, i.e. the predicted probability of a patient being prescribed a drug of interest. The propensity score, ranging between 0 and 1, is used for balancing treated and untreated patients, based on the principle that for groups of individuals with the same propensity score, the distribution of baseline

covariates is the same *(137)*. The aim is then to compare an outcome of interest, such as an AE, in the groups of treated and untreated patients while taking into account propensity scores.

Therefore, there are four common ways in which propensity scores are used to reduce bias: as a basis of stratification, as a covariate to be adjusted, for weighting study subjects/patients differently, or for matching patients *(95, 136, 137, 139)*. The first two relate to the previously discussed methods, so the propensity score can be used to derive strata for stratification, e.g. using the quintiles of the propensity scores as cut-offs for the strata, or the propensity scores can directly be included as a covariate in a regression instead of multiple covariates being included separately *(95)*. In using propensity scores for weighting study subjects, a synthetic sample of patients is created in which each patient is weighed according to the inverse of the propensity scores, which is then used to analyse the outcome of interest *(138)*. In Propensity Score Matching (PSM), groups of treated and untreated patients with similar values of their propensity scores are identified, and the outcome of interest can then be directly compared between the matched groups of patients *(137)*. The matching can be one-to-one, with one matched untreated subject for each treated subject, or many-to-one, with multiple untreated subjects per treated subject, which may in turn be selected as the most similar untreated patients from a larger group of untreated subjects *(95, 137)*. Figure 1.13 represents the concept of propensity score matching in observational studies involving drug treatments.

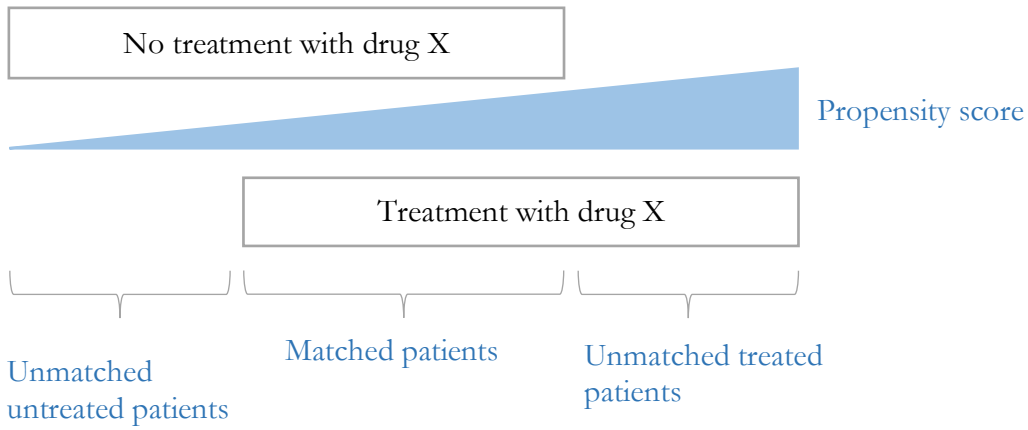

Figure 1.13 Conceptual representation of propensity scores used for matching in an observational study. The outcome of interest is compared in the matched groups of treated and untreated patients

Adapted from *(139)*.

In the context of post-marketing surveillance, propensity scores are the estimated probability that a patient is prescribed a drug based on the patient demographics and pre-existing conditions, i.e. the patient's diagnoses *(98, 138)*. Tatonetti et al. used PSM on the FAERS database to find comparable group of patient reports, of which one group was exposed to the drug of interest and the other was not, before calculating disproportionality statistics using the matched groups of reports *(98)*. The variables included in the propensity score model were co-medications and indications correlated with a drug of interest, and the authors showed that the method reduced false positive associations due to these confounding factors *(98)*. The method also reduced differences in age and sex between the groups used for comparison, even though these variables were not included in the model, which the authors attribute to concomitant drugs and indications being correlated with other patient characteristics for which they are able to serve as a proxy *(98)*. This result is encouraging because these correlated variables may be unmeasured and unknown confounders. The drug-AE dataset from this publication was released as the OFFSIDES dataset *(98)*. However, case-deduplication and standardisation of drug names were not performed in this dataset *(98)*.

An advantage of propensity score methods compared to e.g. stratification is that they can include large numbers of covariates. For example, one study investigated the gastrointestinal (GI) side effects of cyclooxygenase inhibitors in electronic health records, considering between 200 and 500 empirically identified covariates *(140)*. The study showed that the analysis using propensity scores produced results comparable to those of randomised controlled trials of the same effect *(140)*.

1.7.3 The contingency table and related metrics

Once drug-AE and drug-target relationships are identified, the next step towards identifying target-AE relationships is to infer relationships between targets and AEs. Therefore, a measure is needed to study the concordance between these two variables. The relationship between two dichotomous variables can be investigated using a contingency table, also known as a confusion matrix, which counts instances in four possible cells (Table 1.9).

Table 1.9 Example of a contingency table with two categories

		True category	
		Positive	**Negative**
Test result or predicted category	**Positive**	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)

Confusion matrices are frequently applied to evaluate the usefulness of diagnostic tests in medicine *(141)* and the performance of predictive classification models *(142)*. Using the numbers from Table 1.9, the sensitivity and specificity of a test or predictive model can be calculated as follows *(141)*.

$$Sensitivity = \frac{TP}{TP + FN} \tag{1.7.8}$$

$$Specificity = \frac{TN}{FP + TN} \tag{1.7.9}$$

The sensitivity, or true positive rate, indicates the fraction of true cases that have a positive test result or predicted class, whereas the specificity indicates the fraction of true negative cases that have a negative test result or predicted class *(141, 142)*. The sensitivity and specificity represent the accuracy for the true positive and true negative classes respectively *(142)*, but they do not directly indicate the chances that a new sample belongs to the positive or negative class given the test result, which is often a relevant question in practice *(141, 143, 144)*. When applying models to new samples, or doing *in vitro* testing such as in secondary pharmacology screening, the true categories are unknown and only the prediction or test result is available *(142)*. Therefore, predictive values have been suggested as relevant metrics for decision-making in practice *(71, 142, 144)*. The PPV is the proportion of true cases among all positive test results and the negative predictive value is the proportion of true negative cases among all negative test results, and they are calculated as follows *(141, 143)*.

$$Positive\ predictive\ value = \frac{TP}{TP + FP} \tag{1.7.10}$$

$$Negative\ predictive\ value = \frac{TN}{FN + TN} \tag{1.7.11}$$

The studies by Pollard et al. and Krejsa et al. *(69, 71)* discussed earlier in section 1.4.1.3 used PPVs to characterise target-AE relationships, using different thresholds for the IC_{50}.

Predictive values are influenced by the prevalence, i.e. the fraction of individuals in the population with the disease in case of a diagnostic test *(141)*. Higher PPVs are observed with increasing prevalence, and lower PPVs with decreasing prevalence, all else staying the same, because the number of false positive test results increases when testing a population with low disease prevalence *(141, 143)*. On the one hand, the fact that predictive values take into account the prevalence can be useful because it allows the PPV to indicate the probability of a new sample belonging to the positive or negative category *(142)*. On the other hand, it can be a drawback when wishing to compare the predictive values of tests for outcomes with different prevalence, such as different preclinical toxicity findings in animal studies *(81)*, when some events are more common than others. One way to alleviate this is by calculating the value-added PPV, which is the PPV minus the expected prevalence *(141)*. This is related to the idea that an initial estimate of the probability of disease before testing can be set as the prevalence, which is then updated to the PPV after testing *(143)*. Thus, the value-added PPV indicates the additional usefulness of the test in addition to the prevalence as initial estimate, and is calculated as follows *(141, 143)*:

$$Value\text{–}added\ PPV = PPV - \frac{TP + FN}{TP + FP + FN + TN} \tag{1.7.12}$$

Another metric that is independent of the prevalence is the Likelihood Ratio (LR), indicating the likelihood of a certain test result in diseased individuals compared to unaffected individuals, in other words, how many more times likely a positive result is in diseased individuals than in healthy individuals *(141, 145)*. For the likelihood of a positive test result, the LR is calculated as follows *(141, 145)*.

$$LR+ = \frac{sensitivity}{(1 - specificity)} = \frac{(TP/(TP + FN))}{(FP/(FP + TN))} \tag{1.7.13}$$

The LR+ is equivalent to the percentage of positive test results in the positive group over the percentage of positive test results in the negative group *(145)*. For the likelihood of a negative test results, the LR- is calculated as:

$$LR- = \frac{(1 - sensitivity)}{specificity} = \frac{FN/(TP + FN)}{TN/(FP + TN)} \tag{1.7.13}$$

The LR can take on values between 0 and infinity, with high LR+ values occurring when a positive test result is much more frequent among individuals with the disease than unaffected individuals, and the test is considered useful *(145–147)*. Conversely, LR- values close to 0 mean that a negative test result provides evidence for the negative class, and values of LR+ or LR- close to 1 mean that the test does not provide useful information for decision-making, *(145, 146)*. Clark et al. used LRs to study the concordance between preclinical and clinical AEs, using the AE observed in animals as a predictor for the clinical AE *(74, 81)*. In this thesis, the metrics introduced here will be used to evaluate the relationship between the *in vitro* activity and *in vivo* AEs, using the *in vitro* activity as the predictor of the AE. Therefore, continuous data such as IC_{50} values for bioactivity and PRR values for AE associations will be binarized using a threshold to define the positive and negative categories for the contingency table.

1.8 Literature review of studies deriving target-adverse event relationships

Having reviewed relevant data sources and statistical methods, this section will review previous literature that used data analysis of drug-target and drug-AE datasets to derive target-AE relationships.

The study by Krejsa et al., discussed before in section 1.4.1.3, derived associations between compounds' *in vitro* potency at safety targets and ADRs reported on drug labels, as compiled in the Bioprint database *(69)*. Only three examples of such relationships for previously reported target-AE associations were presented: muscarinic receptor M1 and tachycardia; histamine H1 and somnolence; and dopamine D1 receptor and tremor/extrapyramidal symptoms *(69)*. While these associations were already known at the time, Krejsa et al. provided the quantification in terms of PPVs by binning the *in vitro* potencies and reporting the percentage of drugs listing the ADR in each potency bin (Table 1.10) *(69)*. These results show a concentration-response relationship between the *in vitro* potency and PPV for human ADRs *(69)*.

Table 1.10 Approximate PPVs (%), corresponding to the percentage of classified active compounds reported in each potency bin that listed the side effect on the drug label

From the study by Krejsa et al. *(69)*.

IC_{50} potency bin	**Muscarinic M1 receptor – tachycardia**	**Histamine H1 receptor – somnolence**	**Dopamine D1 – tremor /extrapyramidal symptoms**
> 50 μM	25	42	21
10 – 50 μM	39	40	40
1 – 10 μM	42	65	41
0.1 – 1.0 μM	52	80	42
< 0.1 μM	80	90	90

Pollard et al. calculated the sensitivity, specificity, area under the receiver-operator curve (AUROC), and predictive values of *in vitro* hERG inhibition for clinical QT interval prolongation *(71)*. They studied different bioactivity cut-offs (1, 3, 10, 30, 100, 300, and 1000 μM) for activity at hERG, and also considered exposure margins with respect to the unbound C_{max} (3, 10, 30, 100, 300, and 1000-fold), reporting the PPV and NPVs with respect to QT interval prolongation for each of these. The absolute hERG IC_{50} had an AUROC of 0.78 whereas considering the margin between the hERG IC_{50} and the unbound clinical peak plasma concentration resulted in an AUROC of 0.80 *(71)*. The PPVs observed in the study are shown in Table 1.11, showing that the PPV is highest (1.00) when the safety margin between the hERG IC_{50} and clinical C_{max} is less than 10-fold. However, the authors comment that in early drug discovery the clinical C_{max} is not known, so that the PPV of 0.63 is applicable, meaning that drugs with an *in vitro* IC_{50} below 10 μM have a 63% chance of causing clinical QT prolongation *(71)*. While this study is detailed in the quantification of the relationship, it only studied one target-AE combination.

Table 1.11 PPVs for clinical QT prolongation measured in Thorough QT/QTc Studies, when considering two types of thresholds: the absolute hERG IC_{50} or alternatively the safety margin between the hERG IC_{50} and the clinical free plasma concentration. A prevalence of 0.22 was used in the study for the fraction of drugs with QT prolonging effects.

From the study by Pollard et al. *(71)*

Thresholds	**<1**	**<3**	**<10**	**<30**	**<100**	**<300**	**<1000**
As absolute hERG IC_{50} (μM)	1.00	0.54	0.63	0.42	0.37	0.27	0.29
As safety margin between hERG IC_{50} and clinical free C_{max} (exposure multiple)	-	1.00	1.00	0.68	0.54	0.28	0.27

Instead of focusing on the relationships between with single targets and ADRs, Fliri et al. studied the full bioactivity profiles of 872 approved drugs by clustering their *in vitro* bioactivity fingerprints comprised of activity against 92 proteins *(148)*. ADRs from drug labels, derived from both placebo-controlled trials and post-marketing reports, were extracted from the Bioprint database *(69, 148)*. Binary fingerprints of 240 different ADRs were constructed and also clustered *(148)*. Next, the authors compared whether drugs with similar *in vitro* bioactivity profiles also had similar ADR profiles *(148)*. A positive correlation was found between the *in vitro* bioactivity similarities and *in vivo* ADR profiles across different pharmacological drug classes, suggesting that full *in vitro* pharmacological profiles can be used to predict clinical ADRs *(148)*. The authors suggest that *in vitro* biological fingerprints can be used to guide drug discovery towards candidates with fewer ADRs *(148)*.

To investigate the predictive ability of pharmacological fingerprints more explicitly, Bender et al. built multi-class Bayesian models for target prediction using ECFP_4 chemical fingerprints to predict the likelihood of activity at 70 targets on the safety panel at Novartis *(149)*. Using the full predicted bioactivity fingerprint of drugs, the authors then built another set of multi-class models to predict ADRs for drugs, which had been extracted from the World Drug Index *(149)*. Using examples of marketed and withdrawn drugs, the authors showed that the models picked up known target-related ADRs and that the target-ADR relationships ranked highly by the model were supported by literature *(149)*. In a case study of arrhythmia, the authors further showed how extracting underlying features from the ECFP_4 fingerprints can identify features of chemical structures associated with ADRs, i.e. toxicophores *(149)*. Both such toxicophores as well as *in silico* ADR predictions may support virtual screening of drug candidates in drug development *(149)*.

Lounkine et al. also used target predictions but additionally were able to validate some of the predictions *in vitro (63)*. They first used the Similarity Ensemble Approach (SEA) to predict the *in vitro* bioactivity of 656 marketed drugs for a set of 73 proteins included on the Novartis safety pharmacology panel *(63)*. They then tested novel predicted activities *in vitro* and filtered the *in vitro* bioactivities for *in vivo* relevance by comparing to the C_{max} of the drugs *(63)*. Using ADRs from the World Drug Index, Enrichment Factors with χ^2 test q-values were calculated to identify overrepresented target-ADR pairs *(63)*. Based on this, the authors identified novel target-ADR associations and proposed mechanisms behind known drug-ADR pairs *(63)*. One of the main findings was that abdominal pain reported with the use of the synthetic oestrogen chlorotrianisene could be explained by the drug's predicted activity on cyclooxygenase 1 (COX-1), which is known to mediate GI ulceration, as illustrated in Figure 1.14 *(63)*.

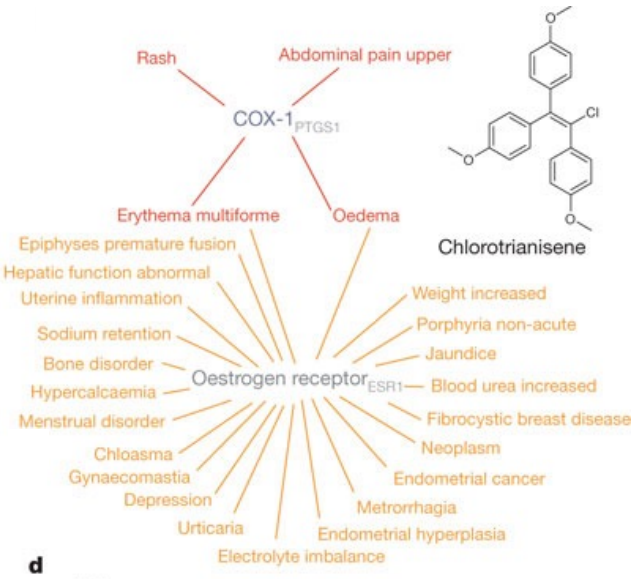

Figure 1.14 Results from the study by Lounkine et al. in which COX-1 was a predicted target of chlorotrianisene. Based on enrichment analysis of multiple drugs, this target was related to abdominal pain, whereas the known target of chlorotrianisene, the oestrogen receptor, was not related to this ADR, leading the authors to suggested the involvement of COX-1 as mechanism behind the abdominal pain.

In the study by Lounkine et al., when multiple targets were associated with the same ADR and drugs were active at these multiple targets, the target with the highest Enrichment Factor was used to suggest a plausible mechanism behind the ADR, thus providing a way to prioritise the numerous associations arising from drug promiscuity *(63)*.

Using experimental instead of predicted target interactions, Kuhn et al. performed a systematic analysis of ADRs for 996 drugs from the SIDER (version 2) database and measured *in vitro* drug-target data for 841 drugs from STITCH3 *(150)*. For 732 out of the 1,428 AEs (51%) analysed, the authors were able to identify overrepresented targets that were statistically linked to the AE based on the Fisher's exact test *(150)*. A total of 262 distinct protein targets were overrepresented, and they provided a plausible causal explanation for the ADR in 70% of the drug-ADR pairs, based on supporting literature evidence *(150)*. Supporting literature included both pharmacological effects as well as effects from mouse knock-out studies. A significant fraction of target-ADR associations could be correlated with identical or related phenotypes observed in mouse knock-out studies of the same target, showing how genetic data can be used to support target-ADR relationships *(150)*. The results from the study included the novel significant association between activation of the serotonin 7 (5-HT7) receptor and hyperesthesia, which is increased pain sensitivity *(150)*. The authors confirmed this effect in mice using a 5-HT7 agonist and antagonist *(150)*. In the methodology used by Kuhn et al., when multiple targets were statistically associated to an AE of which one had been previously reported, the latter target was annotated as the causal explanation and other targets for that event were disregarded *(150)*. While this reduces possible coincidental associations due to compound promiscuity, the method could also limit the possibility of discovering novel targets. A further limitation of the study is that quantifications such as predictive values that can be of value for interpreting secondary pharmacology screening *(71)* were not provided.

Another study by Duran-Frigola and Aloy *(151)* used a similar approach as the studies discussed above but included a greater range of other data types in addition to target interactions. They considered 1,626 ADRs of 992 drugs from SIDER (version 2), relating these to 88 distinct therapeutic drug targets from DrugBank *(151)*. Using the Fisher's exact test, the authors identified 79 overrepresented targets that were statistically associated to at least one of 674 ADRs *(151)*. This means that 41% of ADRs included in the study were significantly associated to one or more protein targets. The study also considered other biological features including protein interactors, biological pathways, and Gene Ontology terms *(151)*. In total 67% of ADRs could be related to a therapeutic

target or another biological feature. Comparing to previous literature, 16% of these biological overrepresented features were found to be previously reported *(151)*. Next, in order to gain an estimate of the comprehensiveness of current molecular information, the authors tried to predict ADRs based on the overrepresented biological features as well as chemical features including small fragments, scaffolds, and ChEBI chemical ontology terms *(151)*. Therefore, the authors built decision tree classifiers allowing multiple features to be used in combination *(151)*. Models for 164 ADRs (14% out of the total 1,162 ADRs with overrepresented features) had an F1 score above 0.5 *(151)*. Out of these, models for 24 ADRs included annotated therapeutic targets as features *(151)*. The authors found that the models relied more often on biological features only (38% of models) versus chemical features only (6%), showing that currently known biological features are important in explaining ADRs *(151)*. A limitation of the study is that only therapeutic drug targets were included as opposed to secondary targets, although protein interactors were included *(151)*. In addition, the associations were not quantified beyond their p-values and non-significant associations were not published, possibly limiting our knowledge of associations that were included in the study but not statistically significant.

Only few studies have investigated targets related to post-marketing AEs and these tend to report on a limited number of associations rather than a large-scale analysis. For example, Maciejewski and co-workers presented three detailed case studies based on FAERS, in which they showed how integrating FAERS data with drug pharmacokinetics can be crucial to discovering the underlying targets and mechanisms behind ADRs *(73)*. In the first case study, the association between 19 drugs targeting the vascular endothelial growth factor receptor (VEGFR) and the ADR hypertension was studied *(73)*. The authors showed that the exposure margin, calculated by taking the ratio of the biochemical IC_{50} over the drug C_{max}, can separate VEGRF inhibitors that are versus those that are not associated with hypertension *(73)*. An exposure margin of 10 nearly perfectly separated the drugs *(73)*. In the second case study, the atypical antipsychotic aripiprazole was compared to risperidone, a drug used in attention deficit hyperactivity disorder *(73)*. Risperidone is associated with the ADRs gynaecomastia and hyperprolactinaemia, whereas aripiprazole is much less so, based on the number of reports in FAERS *(73)*. Although both drugs have similar *in vitro* potencies against the dopamine D2 receptor, the exposure margin for risperidone is narrow (<1) compared to a margin of over 10 for aripiprazole, which can explain the difference in ADR reports *(73)*. This shows how integration with pharmacokinetic data can be important for the interpretation of FAERS data *(73)*. In a similar study of the kinase inhibitor sunitinib, also targeting VEGFR, four targets could be related to the drug's AEs of thrombotic angiopathy and related

effects, based on the drug's bioactivities *(47)*. The target list was narrowed down to one off-target, platelet-derived growth factor, (PDGFR) in addition to the on-target VEGFR (KDR) based on the clinical C_{max} of sunitinib, as shown in Figure 1.15 *(47)*. While these studies demonstrated the approach and potential usefulness of integrating *in vitro* safety pharmacology with pharmacokinetic parameters, this integration was not performed on a systematic scale across all drugs.

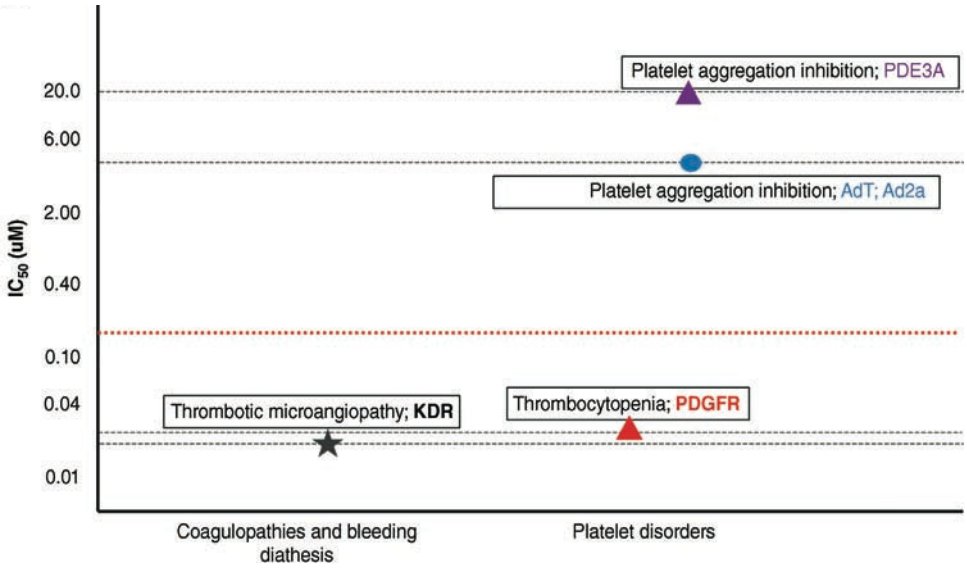

Figure 1.15 Relating targets to reported AEs of sunitinib while taking into account the C_{max} (represented by the dotted orange line). While four targets of sunitinib have known relationships to coagulopathies and platelet disorders, only the potency of sunitinib at VEGFR/KDR and PDGFR is high enough compared to the plasma concentration, to be plausibly related to the ADRs.

Reproduced from *(47)* with permission from John Wiley and Sons © 2015.

Studying the C_{max} of drugs in relation to AEs in greater depth, Redfern et al. identified a link between the safety margin between the *in vitro* IC_{50} at hERG and the unbound drug plasma concentration of drugs, and the risk of TdP. In the study, 100 drugs were classified in different categories according to existing evidence of associated TdP, based on drug withdrawals and the number of AE reports, and it was observed that for almost all drugs without reports of TdP in humans, there was at least a 30-fold separation between the *in vitro* hERG IC_{50} and the unbound effective therapeutic plasma concentration, and this was recommended as a safety margin for drug development *(152)*. This was in contrast to smaller margins for other drugs, such as those withdrawn for TdP, which had a margin between 0.31 and 13 *(152)*.

A different type of study by Sipes et al. also applied the C_{max}-IC_{50} ratio but now on a larger scale *(153)*. Although the study did not specifically focus on ADRs, it highlights the use of above ratio at compound screening stages. The authors used large-scale data from the Tox21 screening initiative, which screened 10,000 environmental, pharmaceutical, consumer and industrial chemicals for activity in biological stress-response assays *(153)*. They estimated the likelihood of the *in vitro* activities being relevant to *in vivo* effects by using predicted C_{max} values, and using the C_{max}-AC_{50} ratio obtained to classify these likelihoods, as illustrated in Table 1.12.

Table 1.12 Boundaries for the likelihood of human *in vivo* interaction in the study by Sipes et al., which used the ratio of the predicted C_{max} and the *in vitro* half-maximal effective concentration AC_{50}

Based on *(153)*.

Classification	**Boundary**
Likely	$\frac{C_{max}}{AC_{50}} \geq 1$
Possible	$1 > \frac{C_{max}}{AC_{50}} \geq 0.1$
Possible with 10x factor	$0.1 > \frac{C_{max}}{AC_{50}} \geq 0.01$

Other studies that used post-marketing data to derive target-AE relationships have mostly been restricted in scope, such as the study by Svensson et al. *(154)*, which focused on structural cardiotoxicity. Using a version of FAERS with standardised drug names by Wang et al. *(155)*, the analysis tried to identify whether drugs with reports of structural cardiotoxicity, identified using a list of relevant clinical conditions (e.g. mitral valve incompetence, left ventricular dysfunction) could be derived from *in vitro* bioactivities in ToxCast *(154)*. Safety signals in FAERS were defined as ROR ≥ 2 and Fisher's exact test p-value ≤ 10^{-4}. The authors ranked associations based on the Mutual Information metric for information dependence, and then investigated the highest-ranking *in vitro* assays against previous literature *(154)*. Based on this, the authors suggested two putative adverse outcome pathways for structural cardiotoxicity involving proteins measured *in vitro* assays: one involving modulation of the translocation protein (TSPO), and another involving reduced expression of Tissue Factor *(154)*. Limitations of the study are that potential confounding factors

were only addressed with a simple filter that excluded any case reports of patients with cardiovascular indications, and only one specific disease area was investigated *(154)*.

Two other studies specifically studied reports of serotonin syndrome in post-marketing data in relation to *in vitro* bioactivities. Serotonin syndrome is a potentially life-threatening ADR presenting clinically as changes in mental status and abnormalities in the neuromuscular and autonomous nervous system, with symptoms including tachycardia, tremor, hyperthermia, and agitation *(14, 156)*. The syndrome is a consequence of excess serotonin in the central and peripheral nervous system resulting from exposure to serotonergic drugs *(156)*. Culbertson et al. used ligand-based target prediction to predict the probability of 71 drugs interacting with different serotonin receptors, the serotonin transporter, the norepinephrine transporter, and muscarinic receptors M1 to M5 *(157)*. Target prediction was used due to lack of a sufficiently complete measured bioactivity matrix *(157)*. The authors then classified the drugs into 7 different categories based on their predicted pharmacology profile and calculated the mean PRR for serotonin syndrome for each of the drug categories using an in-house version of FAERS *(157)*. Serotonin syndrome was defined in the study based on the presence of up to 10 symptoms that together identify the syndrome according to established criteria *(157)*. The drug category with the highest PRR (PRR=1.54 and PRR=1.72 depending on the definition of serotonin syndrome used) contained drugs with simultaneous predicted bioactivities across three target (groups): the serotonin transporter, the norepinephrine transporter, and muscarinic receptors M1 through M5 *(157)*. Since the PRR in this triple-activity group was higher than for drugs acting on the serotonin and norepinephrine transporters only, the authors suggested the potential contribution of muscarinic receptors in the origin of serotonin syndrome *(157)*. Limitations of this study are that the version of FAERS used is not publicly accessible and confounding factors were not addressed.

The other study on serotonin syndrome, by Racz et al., analysed post-marketing data to suggest mechanistic hypotheses for the way second-generation antipsychotics can increase the risk of serotonin syndrome *(158)*. Analysing FAERS through the commercial Molecular Health EFFECT platform and using targets reported for the respective drugs in DrugBank, they calculated the PRR for groups of drugs active at different targets, e.g. drugs active at the sodium-dependent serotonin transporter had a PRR of 8.67 for serotonin syndrome *(158)*. In this case serotonin syndrome reports were used if they directly reported serotonin syndrome. Similarly, they analysed enriched targets among second-generation antipsychotics, and further investigated the overlapping 11 mostly highly ranked targets. Next, the PRR was calculated separately for drugs acting as agonists

and antagonists at each of these targets, which resulted in the greatest signals observed for 5-HT_{1A} receptor agonism, 5-HT_{1A} receptor antagonism, 5-HT_{1B} receptor agonism, 5-HT_{2A} receptor antagonism, 5-HT_{2C} receptor agonism, and α-2A adrenergic receptor antagonism *(158)*. Literature support was found for 5-HT_{1A} agonism and 5-HT_{2A} antagonism as mechanisms potentially leading to serotonin syndrome *(158)*. To investigate confounding factors and potential drug-drug interactions, PRRs were calculated separately for reports listing second-generation antipsychotics in combination with or without serotonin reuptake inhibitors, and this showed that combined use is associated with a disproportionately increased risk of serotonin syndrome, although both classes also had increased risk individually compared to background reports *(158)*. Limitations of the study are that only serotonin syndrome was investigated and a commercial platform was used for the FAERS analysis, but strengths are that care was taken to include functional effects – agonism and antagonism – and confounding factors.

To the best knowledge of the author only one study investigated post-marketing target-AE associations on a systematic scale. The recent study by Ietswaart et al. retrieved AE reports from FAERS and trained Random Forest models to predict these AEs based on the *in vitro* pharmacological profiles of 2,134 marketed drugs against 218 assays measured at Cerep *(159)*. AEs were mapped to 321 HLGTs, which were the dependent variable for the models, and the features were the *in vitro* activities discretised in 3 ranges: highly active (0-3 μM), active (3-30 μM), and inactive (> 30 μM) *(159)*, with inactivity assumed for unmeasured drug-target combinations *(159)*. They mitigated potential biases in FAERS by using the Empirical-Bayes Regression-adjusted Arithmetic Mean, which adjusts for age, sex and reporting year, and by only selecting reports that were submitted by physicians and on which the drug was listed as the primary suspect *(159)*. To discover significant relationships between targets and AEs, the authors inspected the Gini coefficients and identified the target activities with the greatest importance to the models, identifying significant relationships between 51 targets and 221 AEs (MedDRA HLGTs) *(159)*. The results identified previously known relationships such as hERG with various cardiac disorders, but also identified potentially novel AEs for familiar targets, for instance they found that cyclooxygenase-2, hERG, and phosphodiesterase 3 were associated with different kidney disorders, some of which were not previously reported and are suggestions for further study *(159)*. The study considered functional effects for some assays, but did not take into account any pharmacokinetic data *(159)*.

To summarise the studies that focus on target-ADR relationships discussed in this literature review, Table 1.13 shows the main details of the datasets used as well as the number of significant targets reported. It can be seen that no study to date has considered pharmacokinetic data on a global scale using post-marketing data.

Table 1.13. Previous studies using drug-AE and drug-bioactivity datasets to study *in vitro* bioactivities associated with AE

Authors and reference	**Drug-AE dataset**	**Drug-bioactivity or target dataset**	**Association metric**	**Number of targets**	**Number/type of AEs**	**Pharmacokinetics integrated**
Krejsa et al. (2003) *(69)*	Drug labels	Bioprint	PPV per *in vitro* IC_{50} bin	Reported on 3 targets only	Reported on 3 AEs only	
Lounkine et al. (2012) *(63)*	World Drug Index	SEA ligand-based target prediction	Enrichment Factor and q-values	73 targets of Novartis safety panel	317 AEs with significantly associated targets	√, reported for examples only
Kuhn et al. (2013) *(150)*	SIDER version 2	STITCH3	q-values	262 significant protein targets	732 AEs with significantly associated targets	
Duran-Frigola and Aloy (2013) *(151)*	SIDER version 2	Drugbank	p-values	79 significant protein targets	674 AEs with significantly associated targets	
Urban et al. (2015) *(47)*	FAERS (Not specified)	Not specified	Exposure margins (IC_{50}/C_{max})	Case study of VEGFR inhibitor sunitinib (1 target associated)	Thrombotic angiopathy	√, case study only

Maciejewski et al. (2017) *(73)*	FAERS from FDA download	Not specified	Exposure margins (IC_{50}/C_{max})	Case studies of drugs active at VEGFR and dopamine D2 receptor (2 targets associated)	Hypertension, gynaecomastia, hyperprolactinaemia	√, case studies only
Redfern et al. (2003) *(152)*	100 drugs with varying levels of evidence/reports of TdP versus no reports of TdP	hERG IC_{50} from scientific literature	Exposure margins (IC_{50}/ unbound effective therapeutic plasma concentrations)	hERG	TdP	√, unbound therapeutic plasma concentrations
Svensson et al. (2018) *(154)*	FAERS version by Wang et al. *(155)*	ToxCast	Mutual Information	16 significant protein targets and 2 cell-based assays	Structural cardiotoxicity (10 AEs)	
Culbertson et al. (2018) *(157)*	In-house version of FAERS	TargetSearch ligand-based target prediction	PRR per category of drugs with similar bioactivity profiles	21 protein targets studied only	Serotonin syndrome based on 10 potential underlying symptoms	
Racz et al. (2018) *(158)*	Molecular Health EFFECT platform (FAERS)	DrugBank	PRR per group of drugs active at the same single target	Considered 20 proteins most associated, of which 11 studied by agonism/antagonism	Serotonin syndrome as directly reported AE	

Ietswaart et al. (2020) *(159)*	FAERS from OpenFDA	Experimental screening done at Cerep	Feature importance from Random Forest models for AEs	51 targets were significantly related to AEs, out of 218 assays considered	Global analysis of 321 HLGTs, of which 221 significant	

1.9 Aims of the thesis

The aim of this work is to identify and quantify associations between targets and AEs systematically using statistical data analysis of available datasets, while taking into account pharmacokinetics and controlling potential biases in post-marketing data.

Most previous studies using post-marketing data have focused on specific ADRs or have only presented few examples from systematic analyses in their publications, and currently few studies using FAERS data to identify targets associated with AEs have attempted to control potential biases, which are known to substantially affect post-marketing data. Thus, the current work aims to systematically analyse target-AE associations while incorporating existing techniques to control biases in post-marketing data. Therefore, chapter 2, which focuses on the identification of drug-AE associations in FAERS, applies the PSM technique to reduce the effect of confounding factors. The effects of the technique on indication bias are explored using an independent dataset. The resulting dataset of drug-AE relationships is then compared to those derived from clinical trials using the SIDER dataset. Thus, chapter 2 provides datasets of drug-AE relationships for further analysis of associated targets in subsequent chapters.

Although it is widely recommended to take into account pharmacokinetic parameters such as C_{max} in the analysis and interpretation of *in vitro* safety pharmacology results, such integration has only been presented in existing literature as case studies and examples, thus the current work aims to integrate exposure data systematically and explore the impact of doing so. Therefore, chapter 3 is concerned with drug-target links from bioactivity data, providing the other crucial dataset for analysis. In this chapter, measured and predicted *in vitro* bioactivities are integrated with drug plasma concentrations compiled from scientific literature, and it is investigated to what extent this influences the *in vitro* drug-target activity calls.

Given the lack of systematic annotations of safety targets and knowledge of their relevance to drug safety in humans, which are crucial to the interpretation of safety pharmacology screening results, this work aims to contribute to knowledge of target-ADR associations by providing information on *in vitro-in vivo* concordance. Therefore, building on the preparatory work form chapters 2 and 3, chapter 4 comprehensively identifies and quantifies associations between *in vitro* bioactivity at human targets, adjusted for drug plasma concentrations, and AEs from human data, focusing on their predictive values and LRs. This contributes to filling the current gap in knowledge about

targets related to post-marketing AEs, and the identification of potentially novel associations, especially since the amount of post-marketing data is continuously increasing and provides an opportunity to look for emerging associations between targets and AEs. An additional aim is to compare target-ADR associations from clinical trials with those from post-marketing data as this has not been addressed in previous studies.

The results of this work can be used in drug discovery for the interpretation of secondary pharmacology screening results in terms of the human relevance of *in vitro* target activities, as well as for the selection of targets to be used in screening. In addition, knowledge of targets involved in ADRs can improve our mechanistic understanding of adverse drug effects and provide suggestions for the development of predictive computational models.

Figure 1.16 shows the chapter structure of the thesis in support of the above aims.

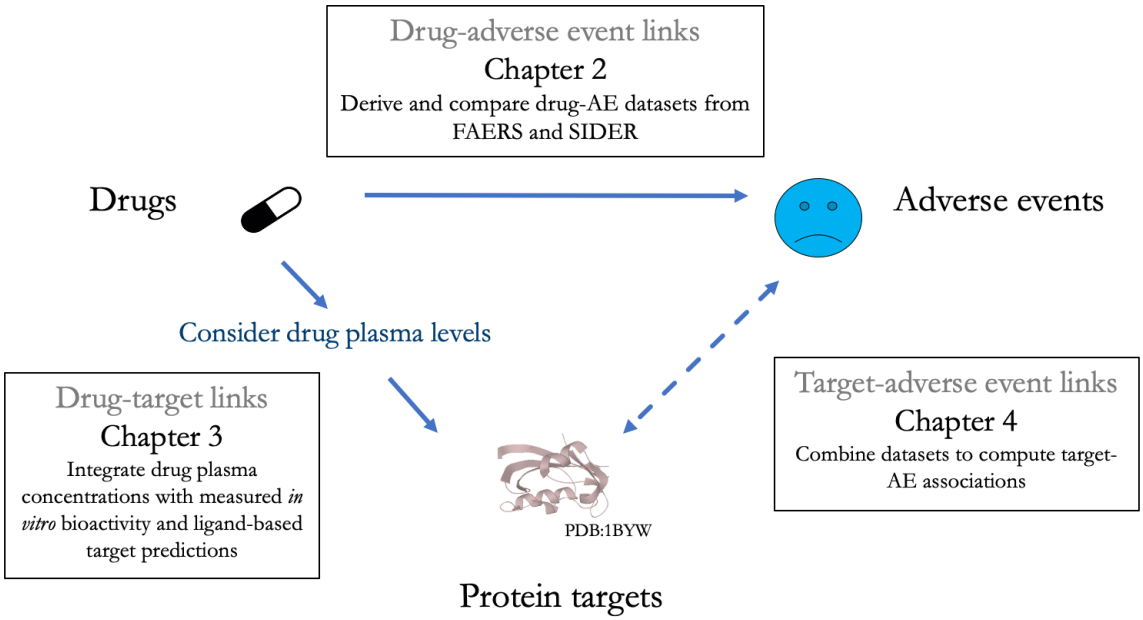

Figure 1.16 Overview of the thesis chapters and how they relate to the main data types analysed

2 Identification of drug-adverse event associations

2.1 Introduction

This chapter will compile two sets of drug-AE relationships, based on FAERS and SIDER, which will be used for further analysis in subsequent chapters. Given the many potential biases that affect post-marketing data, this chapter applies the PSM technique as reported by Tatonetti et al. *(98)* with the aim of reducing confounding factors and improving the reliability of the data. The aim of the PSM procedure is to select a subset of background patients for disproportionality statistics that is as representative as possible of the patients using the drug of interest. Results of implementing this technique on the FAERS AEOLUS dataset, which is larger than the version Tatonetti et al. used *(98)*, are presented.

The curation performed in FAERS AEOLUS consists of (1) deduplication of case reports, (2) standardisation of drug names, and (3) incorporation of FAERS legacy data *(121)*. Duplication of case reports can happen when multiple health professionals submit a report on the same case or when a follow-up report about a patient is submitted *(121, 160)*. Drug names in the original FAERS are recorded as free text and therefore many names for the same drug may be used *(73, 121)*. Both the duplication of reports and the variety of drug names have the potential to affect further statistical analysis on FAERS *(73, 121, 160)*. The legacy data included is FAERS data from the years 2004 – 2012, which follows a different database structure than the current schema, which was introduced in 2012 *(121)*. Thus, the inclusion of the legacy data results in coverage of the years 2004 – 2015.

Because the PSM method is implemented here independently of Tatonetti's code, and on a different database, this chapter starts by reproducing some of the case studies from Tatonetti's work, to generally verify that the implementation works as intended.

Next, the effect of the method on indication bias is investigated using an independent dataset. Tatonetti et al. investigated the effect of PSM on indication bias using a manually constructed dataset of AEs that are caused by the drug indication, but this dataset was not included in the

publication *(98)*. Since drug indications are available from other sources, the effect of PSM on a set of drug indications from the National Library of Medicine is examined in this chapter *(161)*. Indication bias may be more easily observed than other types of bias, and therefore serves as a useful example for verifying the results of implementing the PSM method in this work. The rationale for using an independent dataset is that this provides a reliable, comprehensive source of drug-indication relationships in comparison to the user-submitted drug indications present in FAERS.

Both clinical trials and post-marketing surveillance have limitations, and neither is expected to provide a full profile of possible ADRs. Thus, to complement the post-marketing AEs from FAERS, this chapter also selects ADRs derived from clinical trials from SIDER. Since these datasets will form the basis of most of the analyses in subsequent chapters, this chapter will describe the datasets obtained and present a comparison of drug-AE links derived from FAERS and SIDER.

2.2 Materials and methods

2.2.1 Food and Drug Administration Adverse Event Reporting System (FAERS)

The FAERS AEOLUS database was installed as a MySQL database using the scripts provided by the authors *(121)*. The AEOLUS database contains close to five million unique case reports, 17,710 unique AE terms, 14,063 unique indications, and 3,526 drugs mapped to the RxNorm drug vocabulary *(121)*.

Not all tables from the original FAERS database are included in FAERS AEOLUS. Notably the tables containing the patient age, sex, and reporting country are not included. Thus, these characteristics cannot be analysed using AEOLUS unless linking back to the original FAERS. An overview of the tables in the original FAERS and the information they contain is shown in Figure 2.1, showing which information is retained in FAERS AEOLUS and focused on in the current work. The 'Reactions' are the reported AEs, which are listed in the standard_case_outcome table in AEOLUS.

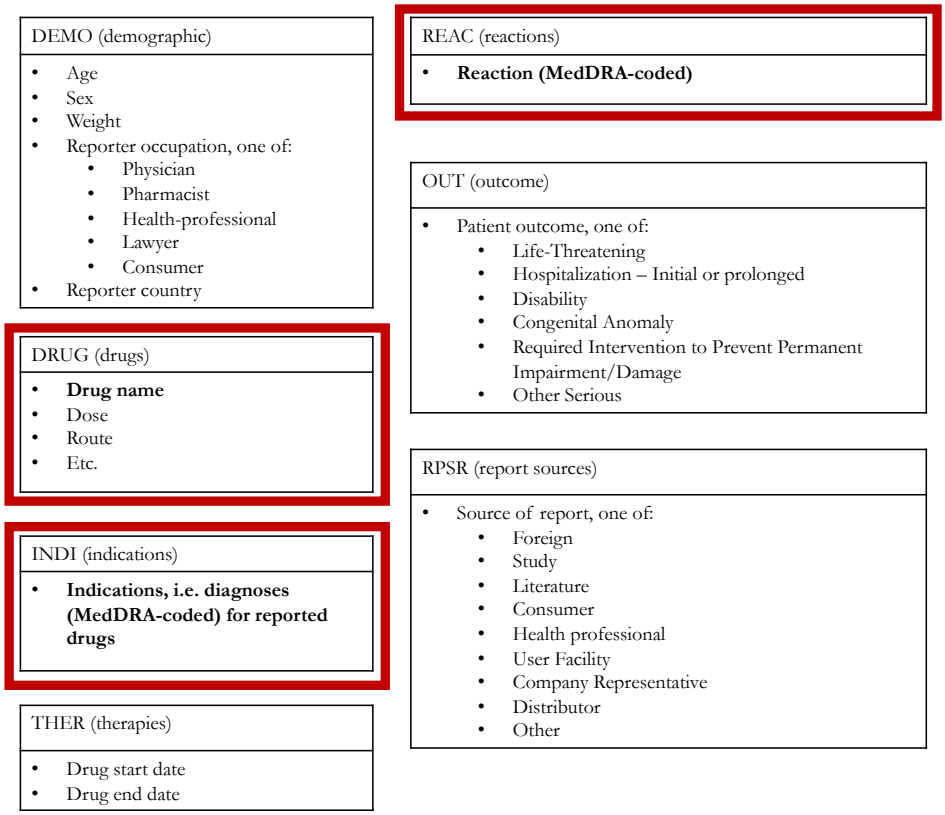

Figure 2.1. Overview of tables in the FDA Adverse Event Reporting System and the main types of information reported in each of the tables. The curated FAERS AEOLUS version only

includes data from the tables highlighted in red. The fields marked in bold are the main fields used in the current work.

2.2.2 Propensity Score Matching (PSM)

The current work closely follows the reported PSM method by Tatonetti et al. *(98)* and all work was done in Python. A separate logistic regression model was implemented per RxNorm drug in AEOLUS using the scikit-klearn (0.21.3) linear_model.LogisticRegression function, to estimate the propensity scores. The model tries to estimate the probability of a patient report listing the drug of interest, i.e. dependent variable, using the independent variables which in this study are dummy variables for the presence of co-reported drugs and indications on the report. Patient reports were identified by the 'primaryid' or 'isr' report identifiers in AEOLUS. To increase the possibility for data-driven discovery, all co-reported drugs were included, regardless of whether they were listed as primary suspect drugs. As discussed in the original publication *(98)*, the number of features in the model must be restricted to prevent a model with thousands of features, thus only drugs and indications associated with the drug of interest were included. The 200 most highly ranked positively associated (Fisher's Odds Ratio > 1) statistically significant (corrected p-value <0.05) co-reported drugs and indications, sorted by their corrected p-values, were identified using the two-sided Fisher's exact test (fishers_exact function) from the SciPy (1.3.2) stats package *(162)* and Benjamini-Hochberg multiple testing correction using the Statsmodels (0.9.0) multipletests function *(163)*. Separate multiple testing correction was performed on the drugs and indications.

Using the propensity score model for each drug, propensity scores were calculated for all patient reports listing a drug of interest, referred to as exposed reports, and up to 100,000 (to limit computational cost) non-exposed background reports sampled randomly from all reports listing at least one of the associated drugs or indications. Next, the range of scores for the exposed reports was divided into twenty equally spaced bins, and for each bin containing reports, up to ten times as many 'control' reports with propensity scores in the same range were selected from the background reports. If the number of control subjects in a certain bin was fewer than the number of exposed reports, the control reports were sampled with replacement. When there were no control reports available, the bin was excluded. At the end of these procedures, a set of reports with the drug of interest and matched background reports was available per RxNorm drug in AEOLUS.

2.2.3 Detection of drug-adverse event associations in FAERS

To be able to investigate the effect of using PSM, drug-AE associations from FAERS were calculated both while using PSM, as described above, and without using PSM as a baseline scenario. In the latter baseline scenario, the two-by-two table for calculation of the PRR compared the AE among all reports listing a drug of interest to all other non-exposed background reports in the database. In contrast, when using PSM, the matched exposed and non-exposed reports were used in the two-by-two table. Thus, instead of using all background reports, a set of reports matched on propensity scores is used as background in case of PSM.

The PRR was calculated as described in section 1.7.1 according to van Puijenbroek et al. *(125)*, on the MedDRA PT in the standard standard_case_outcome table in FAERS AEOLUS, counting unique report identifiers in each cell of the two-by-two table. The corresponding χ^2 statistic was calculated using the SciPy (1.3.2) chi2_contingency function, while setting a minimum expected count of five reports in each cell of the contingency table *(163)*. Significant drug-AE pairs were identified using a previously reported signal detection threshold used in pharmacovigilance of PRR > 2 and $\chi^2 > 4$ *(88)*. The PRR was only calculated if there was at least one report of the AE in combination with the drug of interest and one among the background reports.

2.2.4 Drug indications

For the drugs from FAERS, corresponding drug indications were retrieved from the independent RxClass API (RxNorm vocabulary version 02-Dec-2019) using the RxNorm drug identifiers provided in FAERS AEOLUS. Encoded by the relationships 'may_prevent' and 'may_treat', the RxClass API provides drug indications as Medical Subject Heading (MeSH) terms from the Medication Reference Terminology (MEDRT) source, which is in turn provided by the Veterans Health Administration *(161)*. Next, to map these MeSH terms to MedDRA terms, the MRCONSO.RRF mapping file provided by RxNorm (RxNorm_full_03092019) was used, which provides MedDRA PT and HLTs *(164)*. Thus, where available, this resulted in a set of drug indications in the MedDRA vocabulary per RxNorm identifier. At least one drug indication was available for 85% of the drugs from FAERS. Reported AEs associated with drugs in FAERS based on the PRR analysis were considered to be a drug indication if they had the same HLT as the retrieved drug indications from RxClass for that drug.

2.2.5 Side Effect Resource (SIDER)

The SIDER version 4.1 dataset was obtained from the download files *(87)*. The meddra_all_se file lists side effects for 1,556 drugs, when counting identifiers that include stereochemistry where applicable ("STITCH stereo compound id column"), connected to 4,251 unique side effects (MedDRA PTs, "side effect name" column). The meddra_freq_file lists side effects marked as post-marketing for 624 drugs and additionally, for around 30% of all drug-side effect pairs from the meddra_all_se file, contains information on the occurrence of side effects in clinical studies. A few records from meddra_freq_file that were listed as being incorrect in the SIDER code repository were removed *(165)*. To select the clinical effects only, first all drug-AE combinations from the meddra_all_se file for 928 out of 1,556 drugs were retained because these drugs did not have any post-marketing annotations in the meddra_freq file. Next, for drugs present in both files, drug-AE combinations listed as post-marketing in the meddra_freq file were excluded unless a frequency was available for that same effect, since frequencies can generally only be derived from clinical trials *(90)*. While these steps create the largest possible set of effects from SIDER originating from clinical trials, it is possible that the above filter of post-marketing effects removes some genuine clinical trial effects. This is because some effects could have been observed in both clinical and post-marketing sections, such as nausea for doxorubicin, but it is not currently possible to distinguish these based on the download files from SIDER.

2.2.6 Drug mapping

The drugs in the FAERS and SIDER datasets have different drug identifiers, thus mapping to a common identifier was needed for comparison. Since a crucial part of the subsequent work described in chapter 3 involves retrieving bioactivities from the ChEMBL database, the ChEMBL molregno was used as the common drug identifier. The molregno is suitable because it is stable, unique, and links directly to a unique chemical structure described by its InChI and a connection table in ChEMBL. Furthermore, relationships between salt forms and parent forms of drugs are provided as relationships between molregno and parent_molregnos in ChEMBL. In this work, the aim was to link AEs with the parent drug, thereby grouping different salt forms of the same active ingredient, based on the assumption that drugs exert their main pharmacological action via the active ingredient.

For the drugs from SIDER, the STITCH identifiers ("STITCH stereo compound id column") were taken from the download files and converted to their PubChem identifier by removing the

extra zeros. Unichem is an online service that provides mappings of compound identifiers in different databases based on the molecule InChI Keys *(166)*. Thus, the PubChem identifiers were submitted to UniChem to retrieve ChEMBL molregnos for version 24.1. This step excludes any drugs without an InChI Key such as biological drugs *(166)*. To group different salt forms of the same drug, the parent_molregno was retrieved from a local MySQL installation of ChEMBL version 24.1 *(167)*.

There is no direct link from the drugs in FAERS AEOLUS, which are mapped to the RxNorm vocabulary, to the molecular structure. The only mapping with a link to molecular structures is between RxNorm and Drugbank provided by RxNorm. After retrieving the most up-to-date RxNorm identifiers from the RxNorm API (version: 03-Dec-2018), the RXNCONSO.RRF file provided by RxNorm (RxNorm_full_12032018) was used to find Drugbank identifiers for the RxNorm concepts *(164)*. UniChem, then containing Drugbank version 5.1.1, was used to map to the DrugBank identifiers to molregnos in ChEMBL version 24. As with the SIDER data, salt forms were grouped according to their parent_molregno from the local MySQL version of ChEMBL 24.1. Further queries were also done on the local ChEMBL database; in case the retrieved Drugbank-to-ChEMBL mapping was to an unapproved drug (ChEMBL max_phase < 4), a direct match of the RxNorm compound name to ChEMBL pref_name, compound_name or synonyms (excluding trade name) of an approved drug was preferred, as it is expected to be more accurate due to manual curation in ChEMBL. For any remaining unmapped drugs, the RxNorm compound name, RxNorm synonyms (excluding NDFRT, SNOMED_US, and CVX, taken from the RXNCONSO.RRF file from RxNorm_full_12032018 *(164)*) were checked for direct matches against the ChEMBL pref_name, ChEMBL compound_name, and ChEMBL synonyms. Mappings from this last step were included based on manual inspection of the matches.

2.3 Results

2.3.1 FAERS drug mapping

Out of the 4,245 RxNorm drugs in the FAERS AEOLUS database, 2,764 drugs could be mapped to a ChEMBL identifier. Concepts that could not be mapped include vaccines and substances such as herbal products or chemicals that are not drugs and could not be mapped to a molecular structure or drug identifier in ChEMBL. Out of the 2,764 mapped drugs, the vast majority (n=2,108) are approved drugs in ChEMBL. The non-approved compounds include herbal products and supplements, e.g. menthol, activated charcoal etc., compounds from clinical trials,

e.g. trimebutine, which is annotated as having reached phase 3 in ChEMBL, and other molecules that are not drugs e.g. oxyquinoline and propylene glycol.

2.3.2 Detection of adverse events in FAERS

Using the PRR, drug-AE relationships could be calculated for 1,345 unique RxNorm drug concept identifiers corresponding to 1,323 ChEMBL drugs when not using PSM. When using PSM, this was 1,414 RxNorm drugs corresponding to 1,388 ChEMBL drugs. While the PRR was calculated for each RxNorm drug separately, the results are integrated at the level of parent drugs when preparing the final dataset of drug-AE associations for further analysis (section 2.3.5). The slightly higher number of drugs with PRR calculations in the PSM scenario could be explained by the oversampling of background reports, due to sampling with replacement (see 2.2.2) leading to the minimum expected count of 5 reports in each cell of the 2-by-2 table for the χ^2 test being reached more often (see section 2.2.3).

2.3.3 Propensity Score Matching reduces differences in patient groups to be compared

The PSM procedure should result in matched groups of patients that are more similar in baseline characteristics than the exposed group is to the group of all non-exposed reports. As an example, Figure 2.2 shows such results of the PSM procedure for the drug prochlorperazine, a drug used to control nausea and to treat schizophrenia *(168)*, where this has worked well. The figure compares the listing of the most strongly correlated drugs with prochlorperazine, which could be potential confounding factors, between the exposed and non-exposed reports. For example, among the reports listing prochlorperazine, 26% also list ondansetron, which is also used against nausea *(169)*, but among all non-exposed background reports in the database less than 1% of reports also list ondansetron, resulting in a 25% difference in the use of ondansetron between the groups of reports when not using PSM (Figure 2.2). After using PSM, 28% of the selected exposed reports list ondansetron and similarly 30% of the matched non-exposed reports (Figure 2.2). Thus, differences in concomitant drugs between the groups of reports to be compared have successfully been reduced, providing a better base for subsequently comparing AEs associated with prochlorperazine between the two groups of reports. In this example, PSM is expected to reduce the chance of associating prochlorperazine with AEs that are actually due to ondansetron.

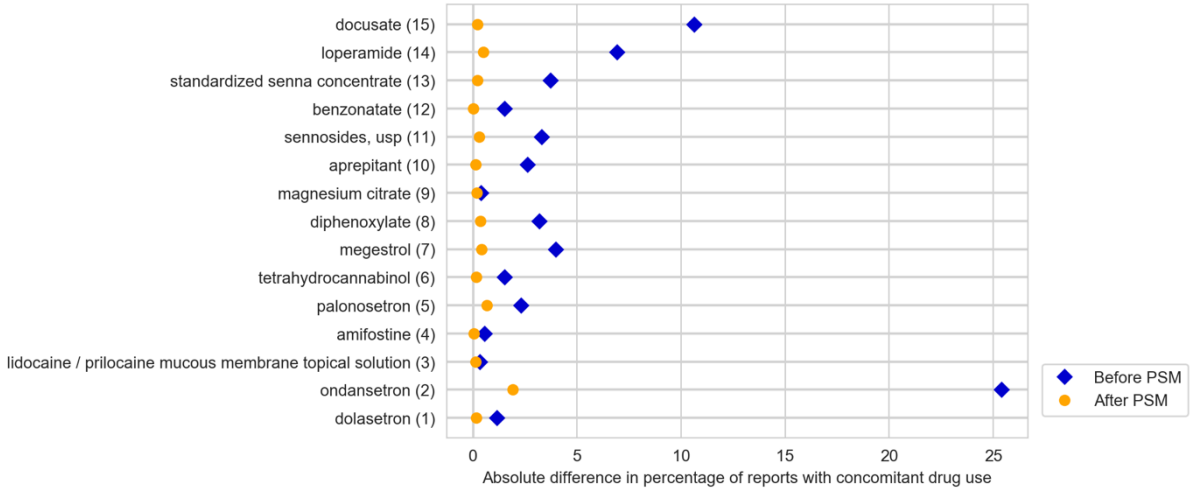

Figure 2.2 Differences in the percentage of exposed versus non-exposed background reports that list one of the 15 most strongly correlated concomitant drugs reported with prochlorperazine in the baseline scenario compared to using PSM. After applying PSM, differences in the listing of concomitant drugs between exposed and non-exposed reports are reduced, providing a better basis for comparing AE reporting rates in the two groups.

Similarly, Figure 2.3 shows the changes for the most strongly correlated drug indications on reports listing prochlorperazine, with particularly the difference in the percentage of reports listing vomiting and nausea being reduced after using PSM. Since both prochlorperazine and ondansetron are used to control nausea *(168, 169)*, the correlated drug indications are consistent with this, and PSM appropriately selected background reports of patients that are more similar to the patients using prochlorperazine in the database.

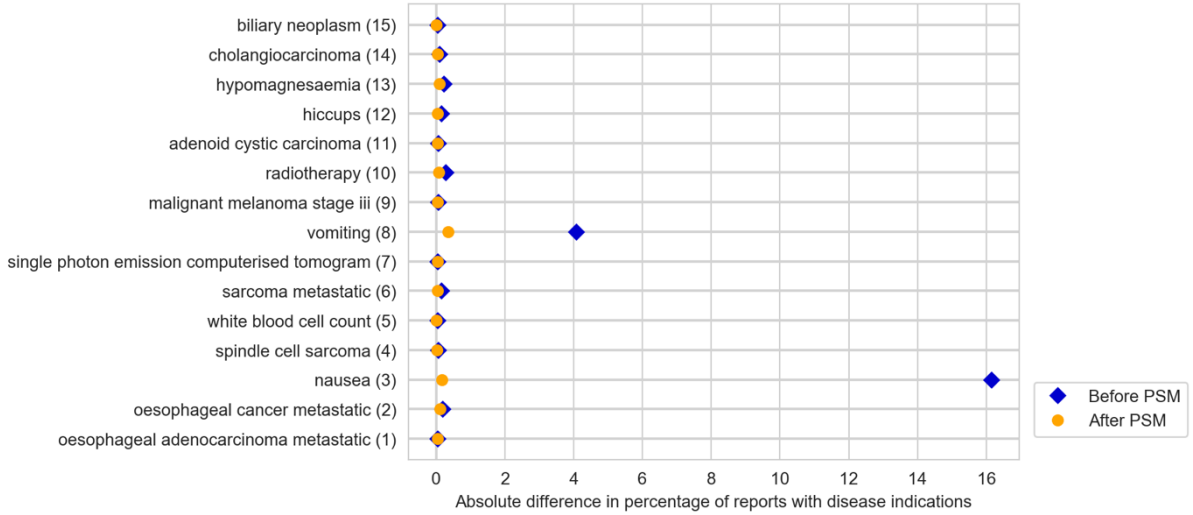

Figure 2.3 Differences in the percentage of exposed versus non-exposed background reports that list one of the 15 most strongly correlated indications reported with prochlorperazine in the baseline scenario compared to using PSM. After applying PSM, differences in the listing of drug indications between exposed and non-exposed reports are reduced.

While the results of prochlorperazine are just one example, similar results were generally observed for other drugs and these cases are representative of the general pattern observed. In some cases, PSM made little difference due to the groups already being fairly balanced, such as for valproate, for which the percentage of exposed and non-exposed reports listing co-reported drug indications differed by less than 1% (Figure 2.4). Occasionally, reducing differences in some attributes comes at the cost of worsening differences in other individual attributes, such as for concomitant drugs with dexfenfluramine (Figure 2.5). Across all the plots discussed, it is also clear that PSM reduces but does not eliminate differences between groups.

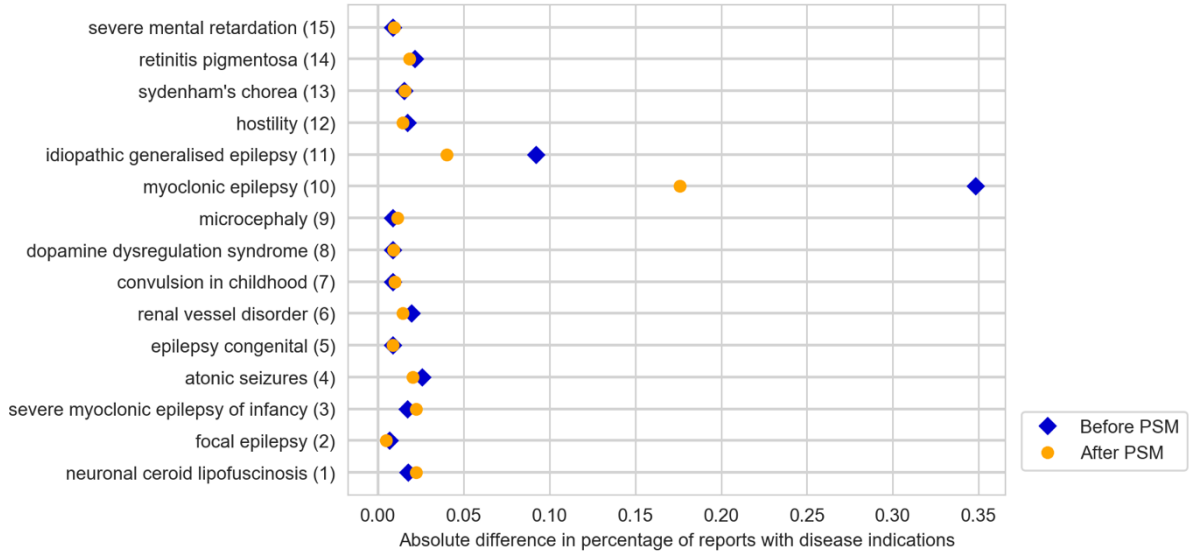

Figure 2.4 Differences in the percentage of exposed versus non-exposed background reports that list one of the 15 most strongly correlated indications reported with valproate in the baseline scenario compared to using PSM. Since the differences are smaller than 1%, in practice PSM makes little difference.

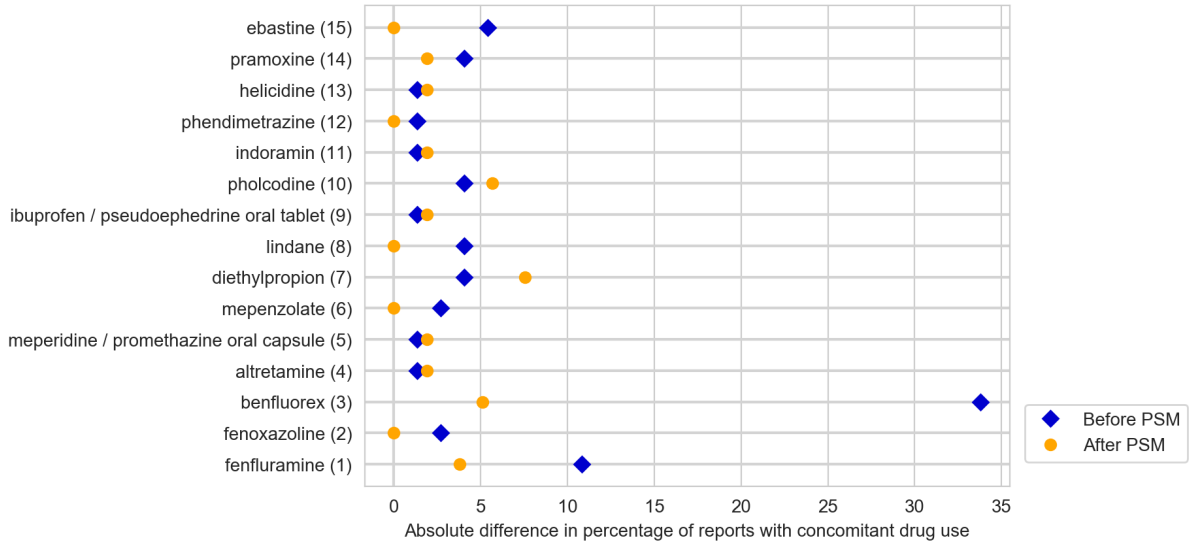

Figure 2.5 Differences in the percentage of exposed versus non-exposed background reports that list one of the 15 most strongly correlated concomitant drugs reported with dexfenfluramine in the baseline scenario compared to using PSM. After applying PSM, most differences in the listing of concomitant drugs are reduced, but some differences have increased, such as for diethylpropion.

While these results are based on visual inspection and only a few representative examples are presented here, the results provide a first indication of PSM having the desired effect. The next question is whether PSM actually reduces false associations with AEs that are actually drug indications, which the next section will consider more systematically.

2.3.4 Propensity Score Matching reduces false associations to drug indications

To examine whether PSM reduces false associations to drug indications, the AEs associated with drugs in the PSM versus the baseline scenario were overlapped with the list of independently retrieved drug indications from RxClass for the corresponding drug. For 1,171 out of 1,388 drugs from the PSM results, drug indications were available, allowing comparison of the PRRs for drug indications in both scenarios. Figure 2.6 shows that PSM preferentially reduces the PRRs of AEs which are in fact drug indications, from a median PRR of 2.5 to 1.2. The PRRs of other AEs, i.e. those not being listed as drug indications in RxClass, are reduced from a median PRR of 1.4 to 1.0. In addition to the reduction in the median, the whole distribution of PRRs is reduced to a greater extent in the case of drug indications (Figure 2.6), e.g. the 75th percentile is reduced from 4.5 to 1.7 for drug indications compared to a change from 2.2 to 1.4 for non-indications. This shows that PSM, as expected, reduces the PRR of confirmed false associations to a greater extent than other associations. These other associations, i.e. non-indications, of which the PRRs are

reduced from a median 1.4 to 1.0 are expected to represent (1) true associations to AEs, but also (2) additional false associations to indications or related symptoms that were not listed in RxClass, and (3) potentially false positive associations to AEs due to confounding factors. It would be difficult to distinguish the three cases in an automated way. It would not be desirable for the PRR of true associations to AEs - the first of the above three cases – to be reduced, but PSM is intended to reduce the PRR in the last two cases. Therefore, the slight reduction in the PRR for non-indications is consistent with a beneficial effect of PSM. Overall, it is concluded that PSM specifically reduces the PRR of confirmed false associations to drug indications.

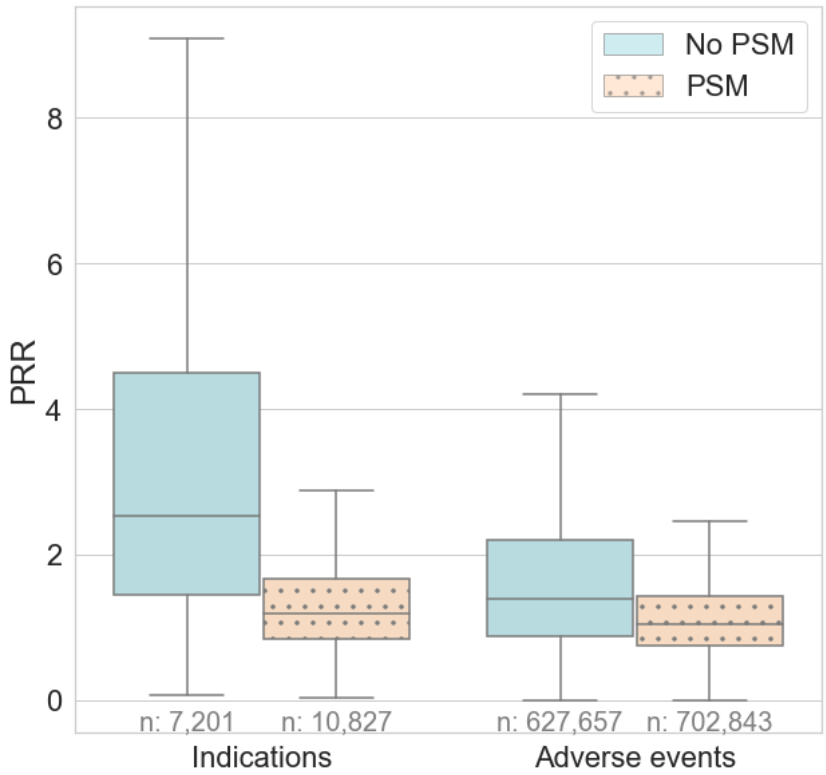

Figure 2.6. Comparison of the PRR of drug-AE associations in the PSM versus No-PSM scenario, split by whether the AE is actually a drug indication. It can be seen that PSM specifically reduces the PRR of drug indications, showing that the procedure reduces indication bias. Boxplot whiskers extend to 1.5 times the interquartile range (IQR).

For cases where a PRR is available for the exact same AE in both the PSM and No-PSM scenario, the change in the PRR is plotted in Figure 2.7. This again shows that for the majority of drug-AE pairs, PSM preferentially reduces the PRRs of drug indications, with the median reduction in the PRR of indications being 1.4, and the median reduction being 0.4 in the case of AEs (Figure 2.7). This supports the conclusion that the PRRs for drug indications are reduced more strongly by the

PSM procedure than PRRs for AEs that are not listed as drug indications in the independent dataset.

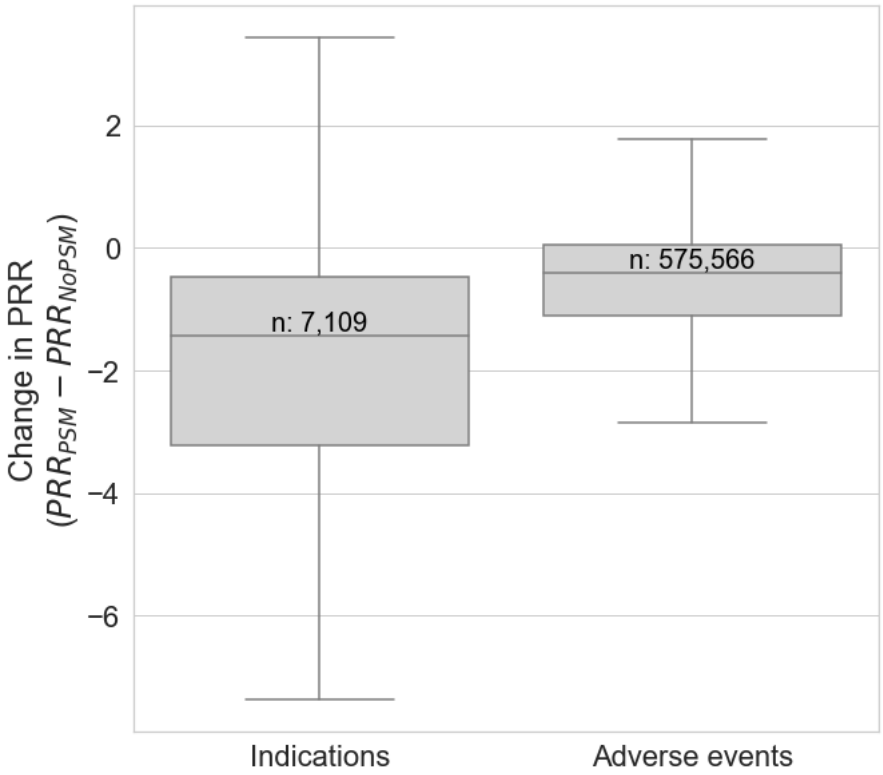

Figure 2.7. Change in PRR of drug-AE associations by PSM across 1,117 drugs. PSM preferentially reduces the PRR of drug-AE associations in which the AE is actually a drug indication. Boxplot whiskers extend to 1.5 times the IQR.

An intuitive example of false associations between drugs and their indications is the association between cholesterol-lowering drugs and hypercholesterolaemia. Hypercholesterolaemia does not represent an AE but an underlying characteristic of the patient population that is prescribed cholesterol-lowering drugs. Tatonetti et al. used this example in a case study *(98)*, which is reprinted in Figure 2.8. Using the same drugs where possible, Figure 2.9 shows that the overall ranges of the PRRs for the same associations are similar in the current study, although the exact numbers differ due to different databases being used. Thus, PSM reduces the PRR of these false associations in the current study with similar magnitude of the effect as in the study of Tatonetti et al. *(98)*. Therefore, it can be concluded that the current implementation of PSM has similar effects as in the original publication of the method in this case study. The implications of the effect by PSM are that the dataset of drug-AE associations derived from using PSM is expected to be more reliable, and contain fewer false positive signals to drug indications.

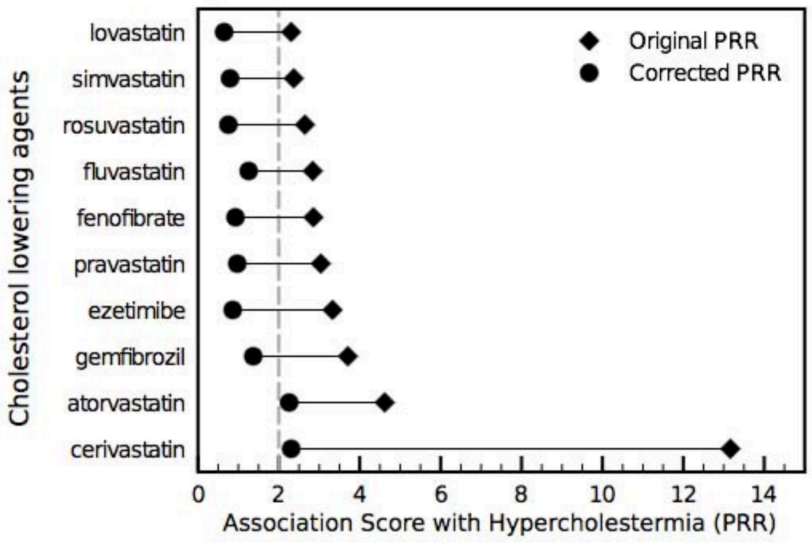

Figure 2.8 Reduction in the PRR for hypercholesterolaemia associated with cholesterol-lowering drugs in the study by Tatonetti et al. Figure from N. P. Tatonetti, P. P. Ye, R. Daneshjou, R. B. Altman, Data-driven prediction of drug effects and interactions, Science Translational Medicine 4, 125ra31-125ra31 (2012). Reprinted with permission from AAAS.

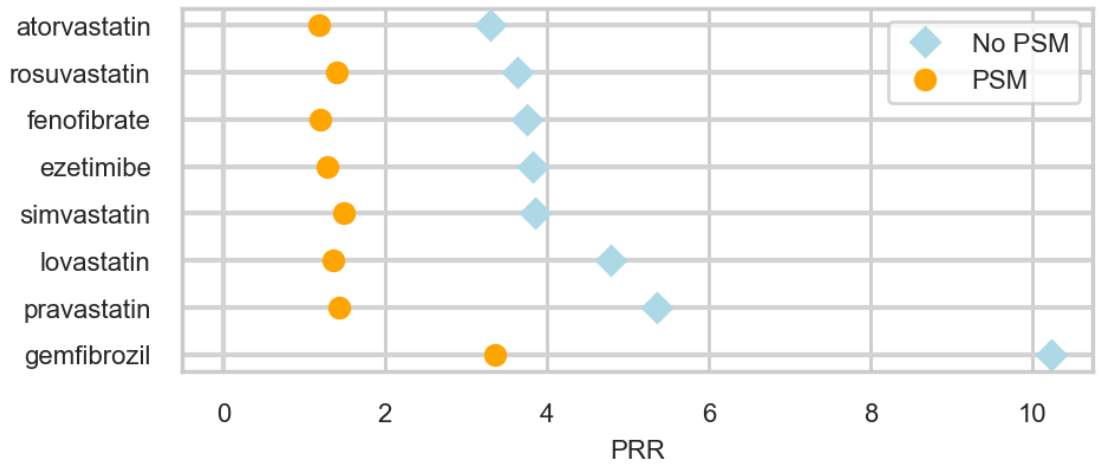

Figure 2.9. Change in PRR for hypercholesterolaemia for the same cholesterol-lowering drugs as in the original case study by PSM in the current study. PSM reduces the PRR of these false associations.

In their other case studies, Tatonetti et al. showed the effects of PSM for specific drugs indicated for diabetes, arrhythmia, and hypercholesterolaemia *(98)*. Now using all the drugs indicated for these diseases in the current dataset of drug indications, Figure 2.10 shows that the implementation of PSM in this work reproduces the desired effects of reducing the PRR for these indications,

often to a PRR below 2, which is a commonly used threshold for the PRR in pharmacovigilance *(88)*. Thus, it can be concluded that the current implementation of PSM has a similar desired effect as the original publication of the method *(98)*. In addition to these cases studies, the previous Figure 2.6 and Figure 2.7 have shown that the effect is observed in the whole dataset, which includes a range of drug indications, and thus is not limited to these case studies.

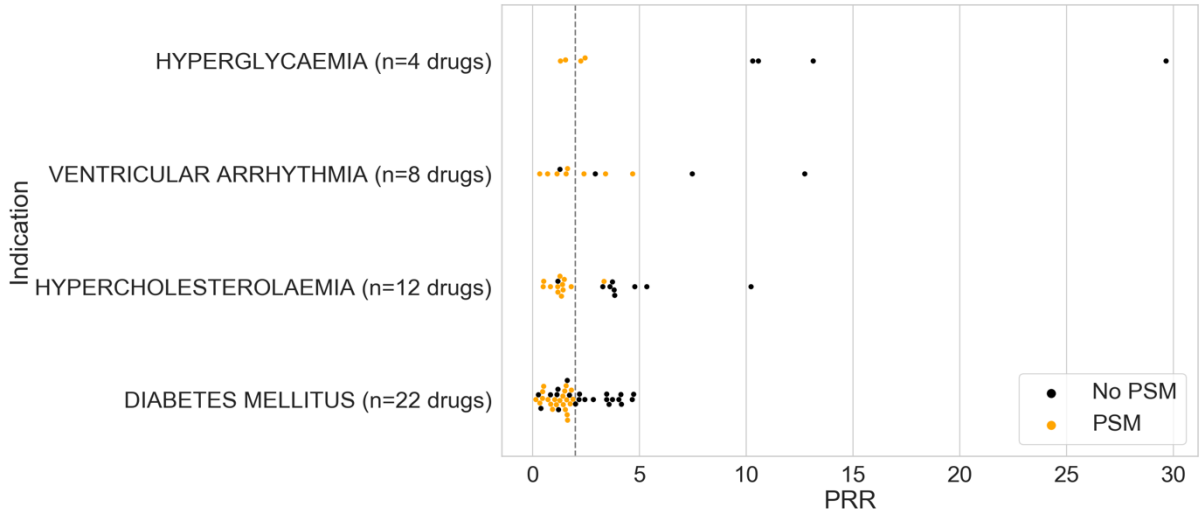

Figure 2.10. PRRs of drug indications in the PSM and No-PSM scenarios, for the drug indications of Tatonetti et al.'s original case studies but using the drugs from the current independent dataset of indications. PSM generally reduces the PRR of false associations to drug indication. The dotted line indicates a threshold of PRR=2.

To investigate in more detail whether PSM behaves differently across different types of drug indications, Figure 2.11 shows the change in PRR per MedDRA SOC of the associated AE. While the magnitude of the effect varies between SOCs, PSM generally changes the PRR in the right direction across SOCs, except for the class 'congenital, familial and genetic disorders', shown by the median of the boxplot for the indications being greater than zero. This suggests that PSM is effective at reducing indication bias across a wide range of drug indications. However, the whiskers of the boxplots extend to changes in PRRs greater than zero, which indicates an increase in the PRR of drug indications falsely associated with some drugs. Thus, it appears that for a minority of datapoints PSM increases rather than dampens the PRR, which is the opposite of the desired effect.

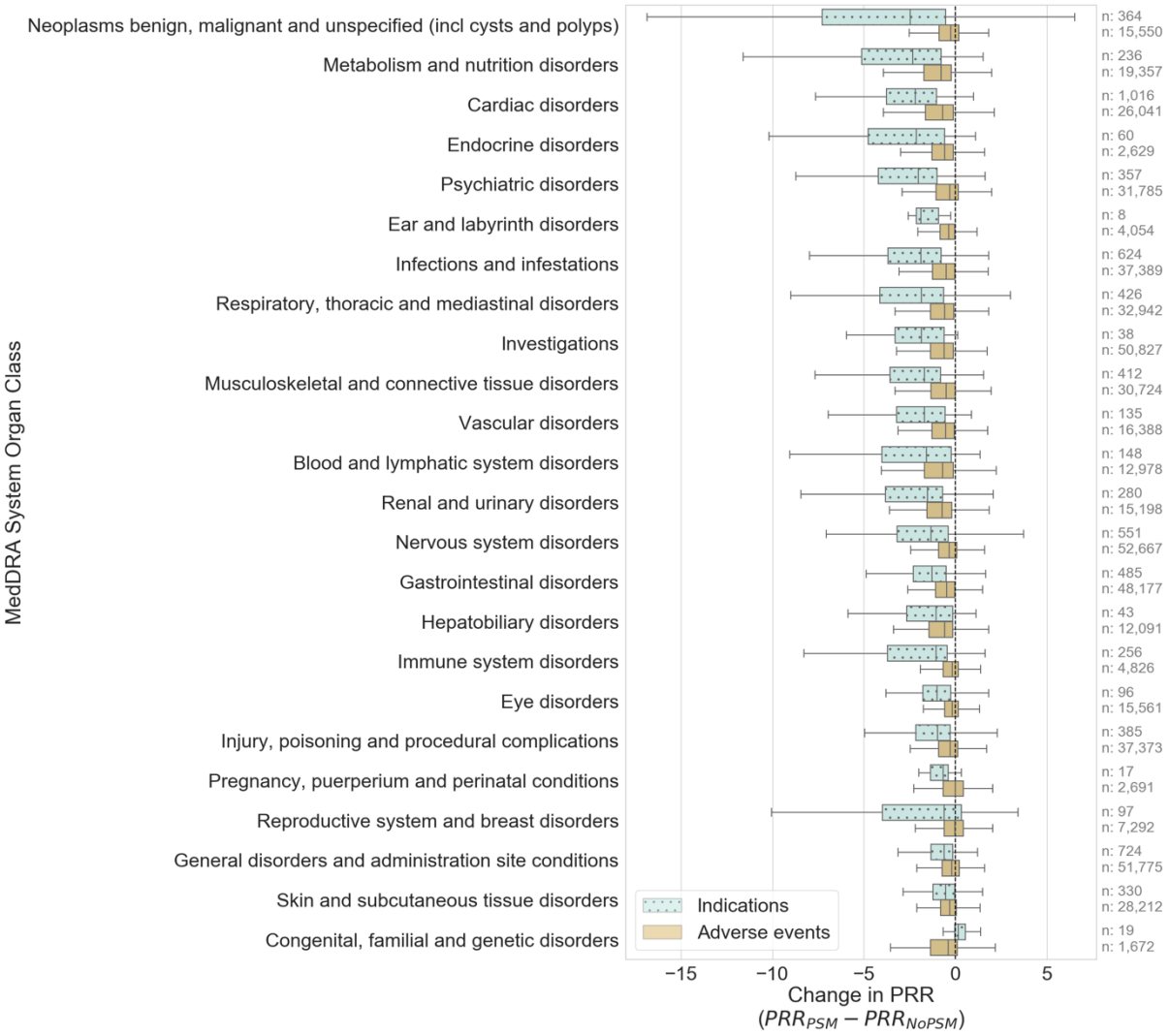

Figure 2.11. Change in PRR of drug-AE associations by PSM per primary MedDRA SOC of the AE (or falsely the drug indication). PSM generally reduces the PRR of drug indications across all SOCs except 'Congenital, familial and genetic disorders', showing that PSM works for a wide range of drug indications. Boxplot whiskers extend to 1.5 times the IQR.

A PRR ≥ 2 and $\chi^2 \geq 4$ is a common significance threshold in pharmacovigilance *(88)*, and Figure 2.6 already showed that for over 75% of drug indications in the PSM scenario, the PRR is below 2. It would be expected that PSM would reduce the overall fraction of statistically significant drug-AE associations that are in fact drug indications. While the median percentage of drug-indication pairs among the significant drug-AE pairs per drug is only slightly reduced from 0.9 to 0.7% (Figure 2.12), the total percentage of drug-indication pairs when adding all drugs together remained the same at 2.3%. Perhaps this is related to the increase in PRR for some drug-indication pairs in the PSM scenario noted from the distributions in Figure 2.11. Figure 2.12 shows that for a minority

of drugs the percentage of drug-indication-pairs among all drug-AE pairs per drug is indeed higher in the PSM scenario. Thus, while PSM appears to have the desired effect overall, there are cases where it is not effective or counterproductive, potentially contributing to masking an effect on the total share of indications among the associated AEs. The fact that PSM also seems to reduce the PRR of non-indications to below the significance threshold could also be compensating for the reduction in associations to indications, resulting in the overall percentage staying the same.

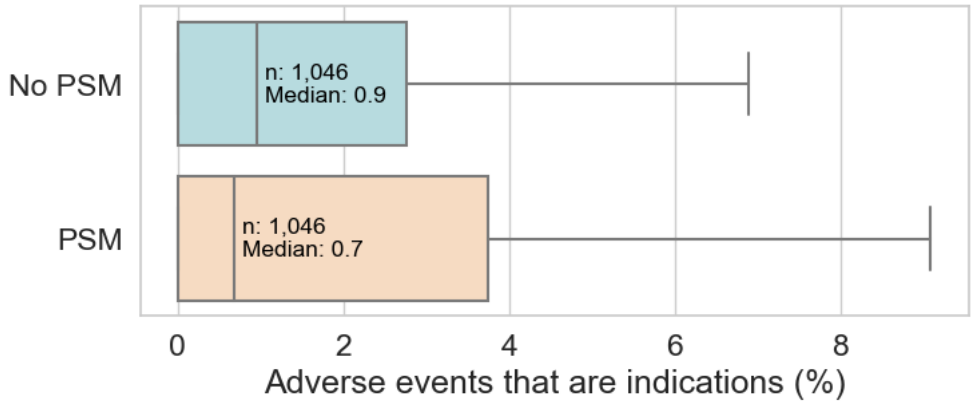

Figure 2.12. Percentage of drug-indication pairs among the significant drug-AE associations per drug for 1,046 drugs (PRR > 2 and χ^2 >4). While the median percentage of indications is reduced, a minority of drugs has a higher percentage of indications in the PSM scenario. Boxplot whiskers extend to 1.5 times the IQR.

2.3.5 Size of drug-adverse event datasets for further analysis

Based on the previous results, the drug-AE associations derived from FAERS using PSM are taken forward for further analysis. Requiring a minimum of 5 drugs per AE results in a dataset of 1,263 ChEMBL parent drugs significantly associated with one or more of 3,365 unique AEs (PTs), comprising a total of 72,200 significant drug-AE associations. This will be referred to as the FAERS dataset of drug-AE associations.

For the SIDER dataset, a total of 1,219 drugs were mapped to molecular structures in ChEMBL. After making the selection of effects derived from clinical trials as described in the methods and excluding side effects with fewer than 5 associated drugs, the clinical trial-focused dataset contains 1,027 ChEMBL parent drugs, which in total are linked to 1,131 unique side effects (PTs). The total number of drug-AE links is 47,324. This will be referred to as the SIDER dataset of drug-AE associations.

The drugs of both datasets will be used to retrieve corresponding bioactivities in chapter 3, and the drug-AE associations which be used to derive target-AE relationships in chapter 4. Therefore, the remainder of this chapter will consider the overall differences between these datasets in terms of drug classes and AE classifications.

2.3.6 Comparison of significant drug-adverse event associations in the FAERS and SIDER datasets

2.3.6.1 Number of significantly associated adverse events per drug

Figure 2.13 shows the distributions of the number of significantly associated AEs per drug in the FAERS and SIDER datasets. The median number of associated AEs is 26 for FAERS and 29 for SIDER, whereas the maximum number of associated events is 896 and 388 respectively. Thus, the FAERS dataset contains a small number of drugs with a high number of significantly associated events. Overall, the distributions of the number of significantly associated AEs per drug are similar between FAERS and SIDER. However, there are large differences between drugs, i.e. some drugs are associated with many AEs at the same time, and others with only few AEs. The 75th percentile of both distributions is just over 60 AEs. Thus, in the majority of cases, drugs are associated with fewer than 60 AEs. The exponential shape of the distribution is in line with another study of FAERS that observed a similar distribution, commenting that 90% of ADRs were related to 40% of the drugs *(73)*. Overall, it can be concluded that the number of associated AEs varies but is most frequently less than 60 AEs.

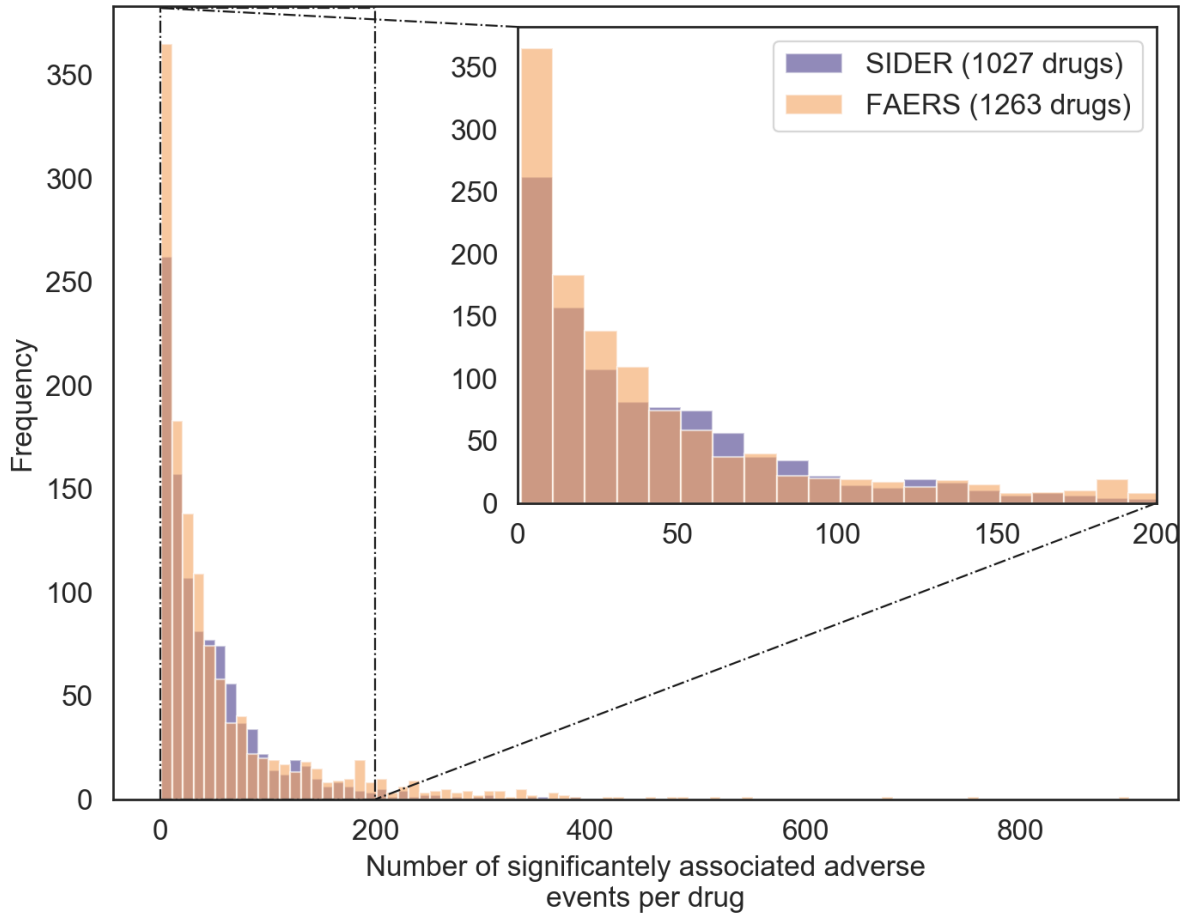

Figure 2.13. Distribution of the number of significantly associated AEs per drug in the FAERS (PRR > 2 and χ^2 > 4) and SIDER datasets

2.3.6.2 Distribution of drugs across Anatomical Therapeutic Chemical Classes

To assess whether any of the datasets is biased towards any therapeutic area, Figure 2.14 shows the Anatomical Therapeutic Chemical (ATC) classes of the drugs in the FAERS and SIDER datasets compared to those of all approved small molecule drugs in ChEMBL 25. The drugs' distribution across ATC classes is largely similar to the set of all small molecule drugs in ChEMBL (Figure 2.14). However, there are some differences, for example, the percentage of drugs used for nervous system indications is 17% higher in the FAERS and SIDER datasets than in the set of all marketed small molecules (Figure 2.14). Nevertheless, both datasets represent all ATC classes of marketed drugs, thus providing a fair representation of therapeutic classes and it can be concluded that neither dataset is strongly biased towards any ATC class.

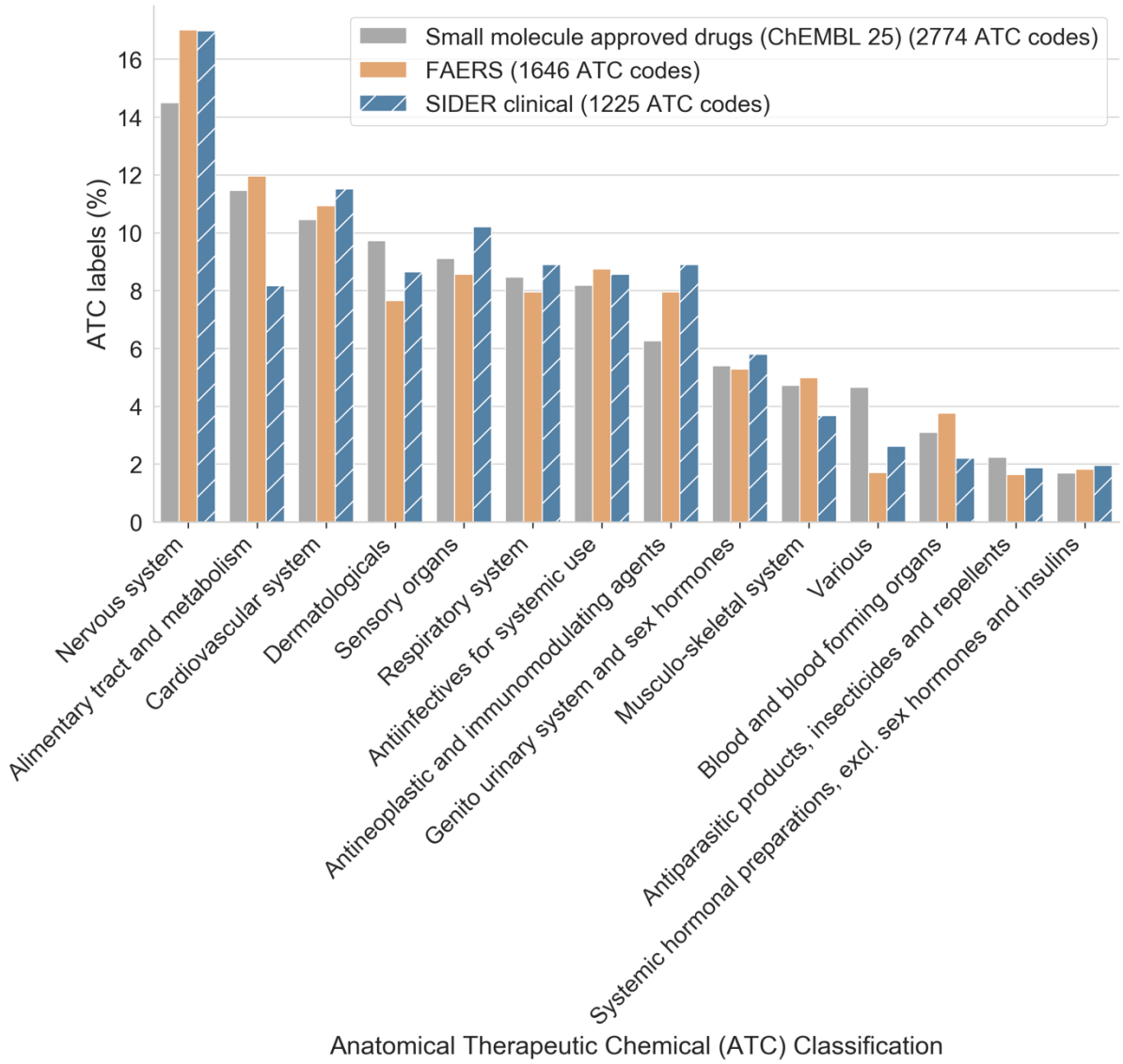

Figure 2.14 Distribution of drugs with significant associations to AEs in the FAERS and SIDER datasets across ATC classes compared to approved small molecule drugs in ChEMBL 25. Multiple ATC labels per drug were included where this occurred. Both datasets appear to be representative of small molecule therapeutic classes.

2.3.6.3 Overlap in drugs and adverse events between datasets

A total of 696 drugs are shared between the FAERS and SIDER datasets, which is ~44% of the union of unique drugs. A remaining ~36% of drugs are only present in the FAERS dataset and a further ~21% of drugs occur in SIDER (Figure 2.15). Thus, both datasets contribute a considerable fraction of drugs not contained in the other.

Regarding the AE terms in the two datasets, ~29% of the unique PTs occur in both datasets, with a remaining 68% and 3% of AEs terms only occurring in FAERS and SIDER respectively (Figure

2.15). This shows that FAERS contains a greater diversity and number of unique AE terms, as the above percentages correspond to 2,349 PTs only occurring in FAERS and 115 only in SIDER.

Comparing the AEs at the HLT level reconciles the discrepancy to some extent, with ~52% of unique HLT terms occurring in both datasets, compared to ~29% of PTs (Figure 2.15). Nevertheless, ~46% of HLTs only occur in the FAERS dataset, versus only ~2% of HLTs that are unique to SIDER (Figure 2.15). Thus, while the FAERS dataset contains more drugs in total, it contains a disproportionately higher number of unique AEs, suggesting a greater diversity of AEs in FAERS. This could be related to the more diverse population from which FAERS reports are derived compared to clinical trials and the focus of post-marketing surveillance on unexpected, rare events.

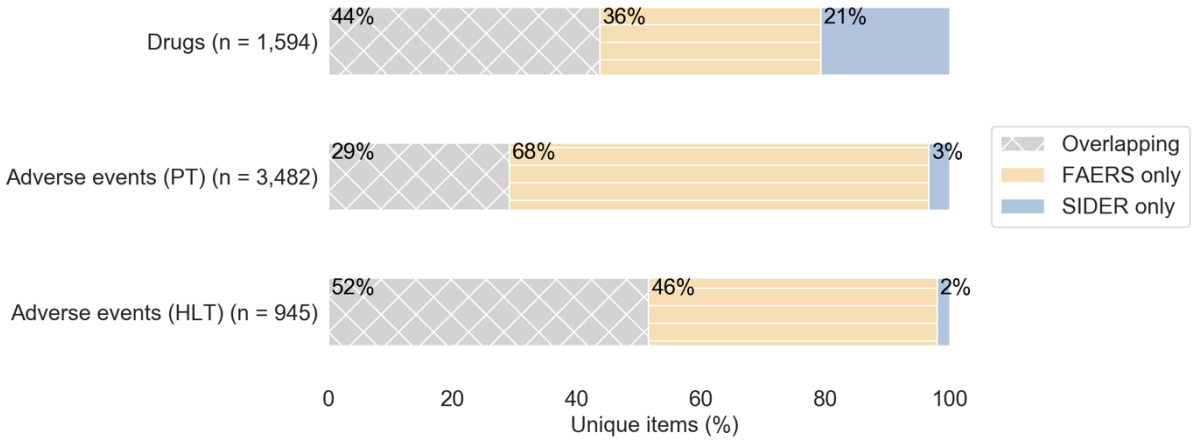

Figure 2.15. Unique drugs and AEs shared between the FAERS and SIDER datasets. The datasets appear complementary because many AEs only occur in either dataset. The FAERS dataset has a large number of AEs not shared with SIDER.

2.3.6.4 Drugs are associated with different adverse events in FAERS and SIDER

To investigate to what extent AEs associated with the same drug are different between FAERS and SIDER, only the AEs associated with the 696 drugs that occur in both datasets were examined. Figure 2.16 shows that for most of these drugs less than 10% (median 1.5%) of all unique AEs listed are shared between the datasets. Therefore, generally, AEs are associated with a given drug only in one of the datasets as opposed to in both, showing that the datasets provide different information.

Furthermore, Figure 2.16 shows FAERS and SIDER contribute a roughly equal share of unique AEs, with the median percentage of AEs being unique to FAERS being 55% and 40% for SIDER.

This means that in most cases, FAERS and SIDER both contribute around half of the unique events per drug. Occasionally, either FAERS or SIDER provides the majority of unique AEs for a given drug, which is indicated by the whiskers of the 'FAERS only' and 'SIDER only' boxplots (Figure 2.16). Both distributions span nearly the entire range, indicating that both FAERS and SIDER occasionally provide nearly all of the unique AEs. Therefore, both datasets provide distinct information for different individual drugs. Overall, these results suggest that FAERS and SIDER provide complementary information.

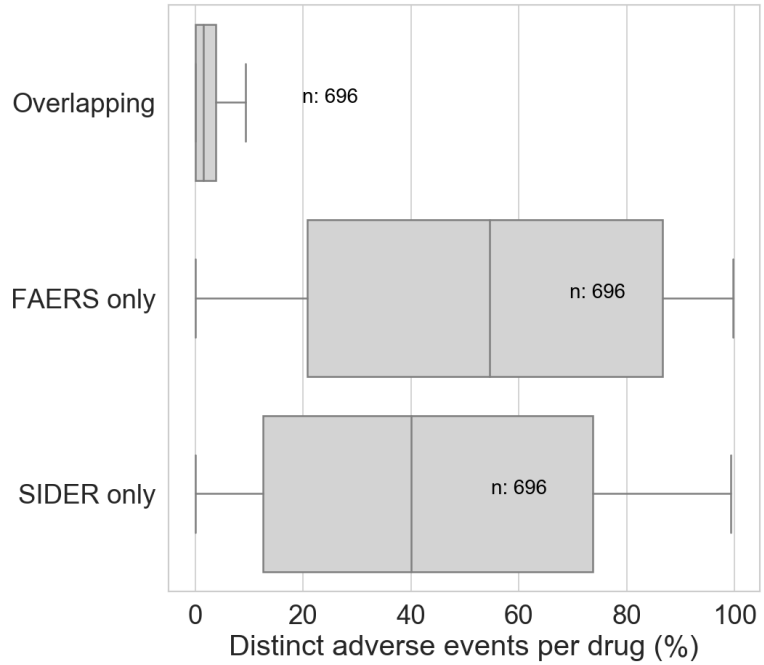

Figure 2.16. Percentage of all unique AEs associated with the a given drug overlapping between the FAERS and SIDER datasets or being unique to either dataset. AEs for the same drug rarely overlap between FAERS and SIDER, with only a median 1.5% of unique AEs overlapping between the datasets. Boxplots show the median with whiskers extending to 1.5 times the IQR.

2.3.6.5 FAERS and SIDER show a different diversity of adverse events across System Organ Classes

Considering the diversity of AEs in the FAERS and SIDER datasets across SOCs, Figure 2.17 shows the distribution of unique AE terms occurring in the datasets, now considering the full dataset again as opposed to overlapping drugs only. 'Investigations' form the largest group of AEs in FAERS, which is a highly diverse SOC that includes events such as electrocardiogram observations and changes in blood pressure, whereas nervous system disorders are the largest group in the SIDER dataset (Figure 2.17).

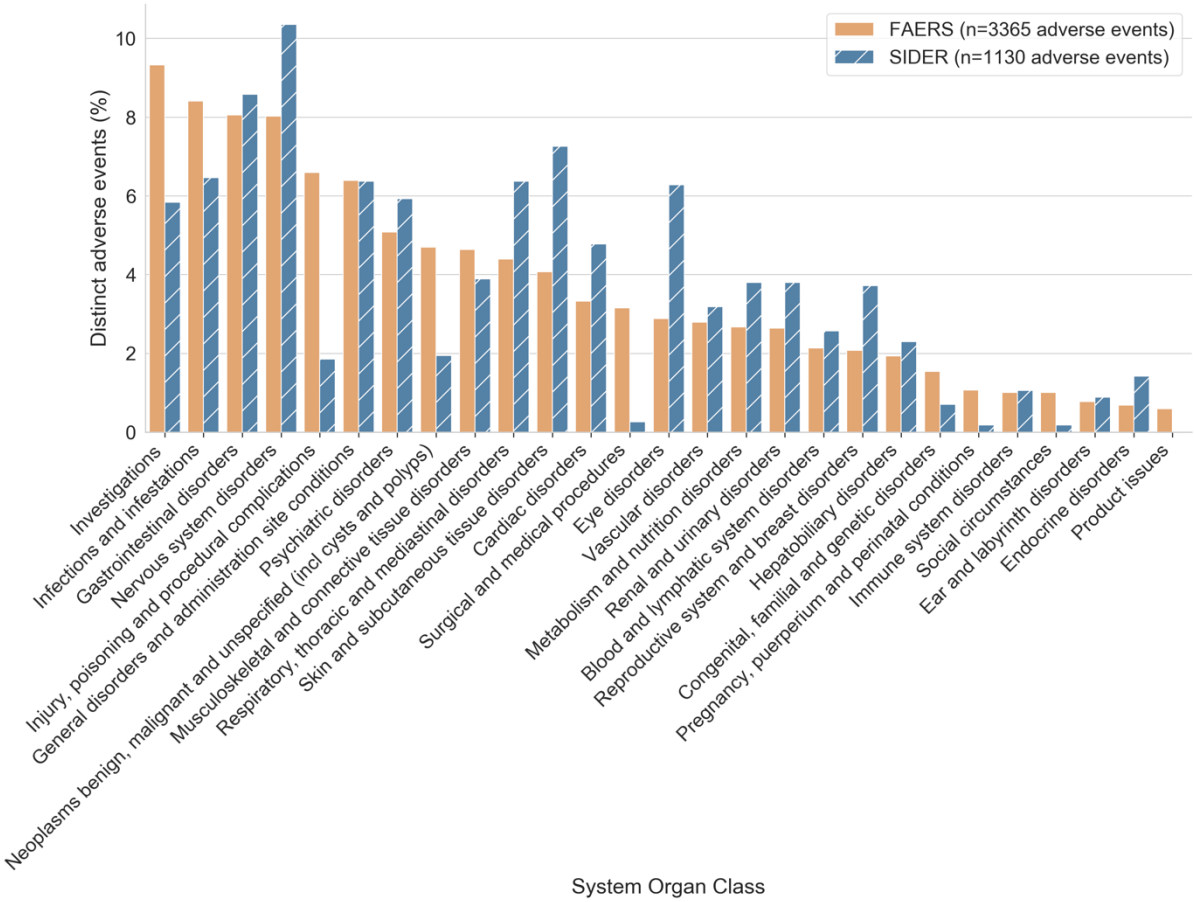

Figure 2.17. Diversity of reported unique AEs across SOC by dataset. Each distinct AE is counted once per dataset, even if associated with multiple drugs. Investigations are the largest group of AEs in FAERS, whereas nervous system disorders are the largest class in SIDER. The largest differences between FAERS and SIDER are seen for 'injury, poisoning and procedural complications', being enriched in FAERS, and eye disorders being enriched in SIDER.

The largest difference between the two datasets is seen for the SOC 'Investigations', which comprises over 9% of unique AEs in FAERS but less than 6% for SIDER (Figure 2.17). Similarly, over 6.5% of AEs in FAERS belong to the SOC 'injury, poisoning and procedural complications', compared to less than 2% of AEs in SIDER. These differences could reflect the fact that FAERS contains reports of accidental and intentional overdoses, medication errors, and product problems *(7)*. The SOC 'surgical and medical procedures' is similarly enriched in FAERS, comprising 3% of unique AEs in FAERS but less than 0.5% in SIDER, which is perhaps also related to the nature of FAERS reporting in terms of accidents and emergencies (Figure 2.17). Another enriched SOC among the FAERS AEs is 'neoplasms benign, malignant and unspecified (incl. cysts and polyps)'. Neoplasms generally take long to develop, thus perhaps this could be related to the longer time scale of drug use for reports submitted to FAERS *(83, 94)*. Lastly, FAERS contains more AEs in 'pregnancy, puerperium and perinatal conditions', which could reflect the longer time frame and

diversity of exposures in the wider population compared to clinical trials *(83, 94)*, e.g. pregnant women are frequently excluded from clinical trials *(170, 171)*. The SOCs mostly highly enriched in SIDER are 'eye disorders', and 'skin and subcutaneous tissue disorders', suggesting that events in these categories are more typically reported during clinical trials than pharmacovigilance reporting. One reason for such differences can be that once drug effects are listed on the drug label, they are less likely to be reported to pharmacovigilance systems *(21)*. Overall, it can be concluded that the AEs in FAERS and SIDER have a different distribution of unique AEs across SOCs, perhaps reflecting their distinct origin.

2.4 Discussion and conclusions

This chapter began with applying PSM on a larger version of FAERS than previously reported and reproducing case studies from the original publication of the method. Using an independent dataset of drug indications, the desired effect of PSM on reducing the statistical association to drug indications was verified, although in a minority of cases PSM appears to increase false associations to drug indications. While indication bias is only one of the potential biases affecting FAERS, it serves as an example for verifying the effects of PSM, and it is expected that other confounding factors are also reduced by the procedure. Previously, Tatonetti et al. showed that when only taking into account drug indications and concomitant drugs in PSM, differences in sex and age were also reduced *(98)*. The result from this chapter not only confirm that the implementation of PSM is working as expected with respect to drug indications, they also strengthen the validation of PSM further to the work of Tatonetti et al. by using an independent dataset of drug indications.

One of the limitations of implementing the PSM procedure in a high-throughput, automated way, is that models are not implemented and evaluated as extensively as would be possible when only studying one drug-AE combination. For example, more sophisticated methods exist for variable selection for propensity score models *(102, 172)* and model evaluation *(173, 174)* than used in this work. These might have the potential to identify cases where PSM is less successful at balancing relevant covariates and reducing the effect of confounding factors, but would require a considerable amount of further work to be implemented on a large scale, so this was beyond the scope of the current work.

An additional limitation is that the PSM method samples non-exposed reports with replacement, meaning that the same non-exposed report can be counted more than once. While this does help to retain as many drugs as possible when suitable background reports are scarce, it might introduce

or amplify imbalances unintendedly. This might be one of the reasons why for a minority of drugs, PSM increased the association to drug indications. Future work could consider whether using one-to-one matching is more reliable. Nevertheless, it is useful that even in the automated way that PSM is implemented here, biases such as indication bias are reduced. These results are in line with another study showing benefits of an automated implementation of a propensity score model *(140)*. Meanwhile, it remains a limitation that propensity score methods do not eliminate biases, but only reduce them.

There are also limitations to the drug-AE relationships compiled from FAERS and SIDER. FAERS is affected by the lack of causal evidence behind AE reports *(7)*, underreporting of AEs, and remaining confounding factors in FAERS, whereas text-mining errors may have affected the SIDER data *(87)*. As a result, some of the drug-AE relationships derived may prove to be incorrect in the future, and at the same time some true drug-AE associations may be missing from the datasets compiled. Nevertheless, this chapter has shown that PSM reduces potential biases in post-marketing data and has helped to create more reliable dataset of drug-AE associations based on FAERS for further study.

Lastly, this chapter compared the drug-AE associations from FAERS to those recorded in SIDER. While the drugs of both datasets represent all therapeutic classes measured by ATC codes, both datasets contain drugs not contained in the other. Moreover, shared drugs are associated with different AEs in the two datasets. The datasets also differ in their distribution of unique AEs across SOCs, with FAERS showing a greater diversity of AEs and potentially reflecting more long-term effects as well as diverse populations. From these results it can be concluded that FAERS and SIDER are complementary, each providing unique AE relationships. The distributions of the datasets presented provide a background against which to interpret the target-AE relationships which will be derived in chapter 4.

3 Adjusting activity calls for measured and predicted *in vitro* bioactivities with drug plasma concentrations

3.1 Introduction

Having prepared the FAERS and SIDER datasets containing drug-AEs links in the previous chapter, the next step towards linking AEs to underlying protein activities is to identify bioactivity data for the drugs from the AE datasets. Therefore, this chapter will use the drugs from the datasets compiled in chapter 2 and introduce their corresponding *in vitro* bioactivities and ligand-based target predictions, and integrate them with drug plasma concentrations. As outlined in the thesis Introduction, it is generally recommended to consider the free drug plasma concentrations alongside *in vitro* drug potencies and affinities to determine the *in vivo* relevance of *in vitro* bioactivities, and this is common practice in later stages of drug development and risk assessment *(19, 20, 57, 66, 69, 73)*. However, existing studies of target-AE associations that use drug plasma concentrations have only reported on a limited number of examples, with no existing study reporting results of a large-scale integration of drug plasma concentrations. On the one hand, analysis of *in vitro* bioactivity data without plasma concentrations represents the early stages of drug discovery *(71)*, since plasma concentrations are determined later in the drug development process. On the other hand, to discover the true relevance of secondary pharmacology target activities for *in vivo* events, the pharmacokinetics must be considered. One likely reason for the lack of systematic studies using drug plasma concentrations is that they are not as easily available in the public domain as *in vitro* bioactivity data. For example, Table 3.1 contrasts the number of drugs that have bioactivity data versus those with C_{max} data in ChEMBL, showing that plasma concentrations are only available for a limited set of drugs. The C_{max} is generally the most abundant parameter to characterise effective drug plasma concentrations in literature, compared to e.g. steady-state concentrations or effective concentration ranges *(36)*.

Table 3.1. Number of parent compounds with bioactivity and C_{max} data in ChEMBL version 24.1 by clinical phase. *In vitro* bioactivities are more abundant than peak plasma concentrations.

Maximum clinical phase of drug	**At least one bioactivity against a human target (n parent compounds)**	**Human C_{max} data (n parent compounds)**
0	735,344	139
1	129	4
2	323	15
3	318	20
4 (approved drugs)	1,578	183

While data availability is limited, no current study has reported to what extent active and inactive labels for drug-target pairs differ when solely based on *in vitro* potency compared to considering both the *in vitro* potency and the drug plasma concentrations. Therefore, in this chapter a set of drug plasma concentrations and plasma protein binding data is compiled from scientific literature and the ChEMBL database. These drug plasma concentrations will be considered alongside the *in vitro* potency or affinity by means of their ratio, similar to the approach introduced in Table 1.12 in the Introduction, to derive an 'adjusted' activity call for drug-target pairs. Thus, this chapter provides intermediate results in terms of the bioactivity datasets, and the way classified active and inactive datapoints are affected by the additional consideration of plasma concentrations. This data provides further background to chapter 4, which will analyse how using the adjusted activity calls affects the relevance of selected targets to human AEs.

3.2 Materials and methods

3.2.1 Extraction of *in vitro* bioactivity data

For the drugs from the FAERS and SIDER datasets mapped to ChEMBL parent compounds, ChEMBL version 25 was queried using a local MySQL installation. Bioactivities against any human single protein or protein complex were retrieved, regardless of these being primary or secondary drug targets. In the case of protein complexes, bioactivities at each of the constituent single proteins listed in the target_components table were listed. The activity types included were 'IC50', 'EC50', 'XC50', 'AC50', 'Ki', 'Kd' and 'Potency' for assay types 'B' (binding) and 'F' (functional) with standard_flag=1. Using the confidence scores of 7 (protein complexes) and 9 (single proteins) ensures the highest level of confidence in target assignment *(107)*. Inactive data was retrieved by extracting records containing any of the following in the activity_comment: 'Not active', 'inactive', 'No inhibition' and allowing standard_flag=0 for these records. The median reported pXC_{50} was computed per parent drug for further analysis and in the case of conflicting active and inactive assignments for the same drug, the active assignment was taken forward.

3.2.2 Protein target prediction

To supplement the measured bioactivity data, the probability of *in vitro* activity at single human proteins was predicted using the ligand-based target prediction tool PIDGIN version 3 *(114, 116)* as follows: compounds were standardised using the e-Tox compound standardiser *(175)*, and the separate classification models available in PIDGIN were used to obtain predictions for the activity below the thresholds of 0.1, 1, 10 and 100 μM *(116)*. The applicability domain threshold was set at a minimum of 0.7 and only models with a minimum Precision-Recall Area Under the Curve of 0.7 during time-series split cross-validation were included *(116)*. Overall, experimental bioactivity values were used where available, with predictions obtained using the settings described above being added to the compound-target bioactivity matrix, but only if at least one measured bioactivity was available for the target, to ensure that subsequent analyses are based on a combination of measured data with supplemented predictions instead of solely on predictions, which would be associated with greater error.

3.2.3 Drug plasma concentrations

Plasma concentration data were compiled from a variety of sources. The upper therapeutic plasma concentrations listed in the supplementary data of the publication by Schulz et al. *(176)* were extracted. In addition, human plasma concentrations (C_{max}) were retrieved from ChEMBL 24.1, allowing the tissues plasma, blood and serum in case the assay_tissue field was filled, and additionally accepting records where the assay_tissue field was empty, thus assuming these are plasma/blood measurements. Furthermore, the standard_flag=1 was used and records with unusual values (data_validity_comment= 'Outside typical range') were excluded, which in case of plasma concentrations enforces a normal range between 1 nM and 10 mM or alternatively between 0.5 ng/ml and 5 mg/ml according to the activity_stds_lookup table in ChEMBL. Values referring to metabolites in either of these sources were excluded based on the mentioning of 'metabolite' or similar in the assay description. Molar concentrations were calculated using the molecular weight of the parent drug recorded in ChEMBL. Where multiple plasma concentrations were available for the same drug and the standard deviation of the pMolar concentration exceeded 1, values were checked against the original publication and corrected where necessary. These corrections affected 8 datapoints from the ChEMBL database. Data for the f_u and PPB of drugs was retrieved from ChEMBL 24.1 and from several publications *(59, 152, 177, 178)*. Where a range was provided, the average of the two values was calculated. For the f_u and PPB extracted from publications, drugs were mapped using the InChI Key and alternatively drug names provided in the publication via direct matching to the pref_name, synonyms, or compound_name in ChEMBL. The median f_u was calculated using the f_u and PPB data, and then multiplied by the total plasma concentrations to derive the unbound plasma concentration. The median total or calculated unbound plasma concentrations were calculated for further analysis.

3.2.4 Activity calls using a constant bioactivity cut-off

To be able to study the effect of integrating bioactivities with plasma concentrations, a baseline dataset was created using a constant cut-off on the pChEMBL value without consideration of the plasma concentrations. Therefore, drugs were assigned as active if the pChEMBL value for the drug-target pair was ≥ 6 ($XC_{50} \leq 1$ μM), else the drug was assigned inactive. This constitutes the activity call using the constant bioactivity cut-off. This threshold was chosen because it is a commonly used *in vitro* bioactivity threshold *(40, 65, 67)*. This dataset will be referred to as the 'constant cut-off dataset'.

3.2.5 Integration of bioactivities and drug plasma concentrations to derive adjusted activity calls

3.2.5.1 Measured bioactivities

The bioactivity measurements and the plasma concentrations were compared to assign an active or inactive label to the drug-target pair, which is the 'adjusted' activity call for that drug-target pair. Therefore, drugs were assigned as active on a given target if the drug plasma concentration exceeded the *in vitro* bioactivity concentration, or was up to 10 times lower (1 log unit) than the *in vitro* concentration. The leeway of 1 log unit is aimed at maximising the potential signal given experimental variability of bioactivity measurements, which may be around half a log unit *(44, 179)*, and considerable inter-individual variability in pharmacokinetics *(176)*. If the plasma concentration was more than 10-fold lower than the measured *in vitro* concentration, the drug-target pair was classified as inactive. Additionally, inactive calls were made based on the activity_comment conveying inactivity as described in section 3.2.1 above. The conditions for the adjusted activity calls are summarised in Table 3.2. Assuming that plasma concentrations correspond to concentrations in target tissues, the adjusted activity calls provide an approximation of whether a target will be modulated sufficiently to obtain an effect.

Table 3.2 Defining the adjusted activity call based on the integration of the *in vitro* bioactivity and drug plasma concentration

Adjusted activity call (label)	**Conditions**	**Definition**
Active	Drug plasma concentration exceeds measured *in vitro* XC_{50}, or	$\frac{C_{max}}{XC_{50}} > 1$
	Drug plasma concentration is up to 10-fold lower than measured *in vitro* XC_{50}	$0.1 < \frac{C_{max}}{XC_{50}} < 1$
Inactive	Drug plasma concentration is more than 10-fold lower than *in vitro* XC_{50}, or	$\frac{C_{max}}{XC_{50}} < 0.1$
	Measured i*n vitro* bioactivity annotated as inactive	'Not active', 'inactive', 'No inhibition' in activity_comment

3.2.5.2 Predicted bioactivities

Depending on whether drugs were in the applicability domain of the target prediction models, probabilities between 0 and 1 were available at all or some of the activity thresholds (0.1, 1, 10 and 100 μM) for each drug-target pair. When predictions were available, a probability below 0.4 was considered inactive and above 0.6 active for each threshold. The lowest predicted active threshold was compared to the plasma concentration, in the same way as for the measured bioactivity data, to derive adjusted activity calls. This is shown for potential scenarios in Table 3.3. For a low percentage of drug-target pairs (<0.01%), the available predictions followed a counterintuitive concentration-response curve, which arises from the models at different thresholds being built on different training sets, and these records were excluded. If all available predictions were inactive, the datapoint was included as inactive if the plasma concentration was lower than 10 times (1 log unit) the highest predicted inactive threshold otherwise the data point was not used (Table 3.3). However, if a compound was in the applicability domain of the models for each threshold, and consistently predicted to be inactive, the datapoint was included as inactive, because the negative predictions span a wide range of concentrations (Table 3.3).

Table 3.3 Target predictions were integrated with plasma concentrations depending on the pattern of active and inactive predictions across the four activity thresholds, as illustrated with potential scenarios.

Activity threshold (μM)				**Description**	**Decision/Adjusted activity call**
0.1	**1**	**10**	**100**		
Active	Active	Inactive	Inactive	Counterintuitive concentration-response curve	Excluded
Inactive	Active	Active	Active	Both active and inactive predictions	Active if plasma concentration exceeds the lowest predicted active concentration (1 μM in example), or is up to 10 times lower (1 log unit). Else inactive.

No prediction available	No prediction available	Active	Active	All active predictions	Active if plasma concentration exceeds the lowest predicted active concentration (10 μM), or is up to 10 times lower (1 log unit). Else excluded (not enough data).
Inactive	Inactive	Inactive	No prediction available	All inactive predictions	Inactive if plasma concentration is lower than 10 times (1 log unit) the highest predicted inactive threshold (10 μM). Else excluded (not enough data).
Inactive	Inactive	Inactive	Inactive	All inactive predictions; compound in applicability domain for all thresholds	Included as inactive even without plasma concentration being available (predicted inactive at wide range of thresholds)

3.2.6 Filter for minimum number of actives

Previous studies *(150, 151)* filtered bioactivity data to have at least 5 drugs classified as active at a given protein, to prevent any further analysis being based on very limited data. Thus, this filter is also applied on the data here after deriving the baseline and adjusted activity calls.

3.3 Results

3.3.1 Origin and size of retrieved bioactivity datasets

The initial dataset of measured *in vitro* bioactivity data for human targets, using the drugs from the AE datasets, was extracted from ChEMBL and contained 113,710 datapoints. Table 3.4 lists the original sources of this data as listed in ChEMBL, showing that DrugMatrix, scientific literature

and PubChem are the main original sources of measured bioactivity data. This is encouraging because data from DrugMatrix and PubChem are derived from screening and thus are full matrices *(105, 107)*. Furthermore, it is clear that DrugMatrix is a larger source of bioactivity data for approved drugs in the AE datasets in this study than scientific literature, on which ChEMBL is primarily based.

Table 3.4 Sources of ChEMBL bioactivity data for the drugs from the AE datasets. DrugMatrix is the largest source of measured bioactivity datapoints for drugs in this study.

Source of ChEMBL bioactivity data	**Datapoints (%)**
DrugMatrix	82.16
Scientific Literature	10.42
PubChem BioAssays	7.022
TP-search Transporter Database	0.291
BindingDB Database	0.062
Patent Bioactivity Data	0.041
K4DD Project	0.004
Deposited Supplementary Bioactivity Data	0.001

To examine the number of drugs and targets present in the dataset, Table 3.5 shows that the experimental bioactivities include 1,147 unique parent drugs and span 1,044 unique targets. However, only 4.8% of the full matrix is filled with datapoints with the rest of the cells being empty, i.e. no data available (Table 3.5). After inserting the predictions into the matrix of measured datapoints, the matrix density of filled datapoints increases to 26.6% of the full matrix. Overall, predicted datapoints comprise 86.2% of all datapoints after combining experimental and predicted data (Table 3.5). This shows that using ligand-based target predictions increases the number of datapoints for further analysis considerably.

In terms of the percentage of active versus inactive datapoints, based on the constant cut-off of pChEMBL value ≥ 6, a total of 6.7% of experimental datapoints are classified as active (Table 3.5). Among predicted datapoints, a lower total of 0.4% of datapoints is predicted active at any of the prediction thresholds, resulting in an overall percentage of 1.3% of datapoints classified as active (pChEMBL value ≥ 6 or predicted active) in the combined matrix of experimental and predicted bioactivities (Table 3.5). Thus, it can be concluded that target predictions primarily add negative (inactive) datapoints when using the absolute cut-off. Furthermore, it is concluded that predictions comprise the majority (86.2%) of all drug-target datapoints, and that in the overall dataset of combined experimental and predicted values, 1.3% of datapoints is classified as active. The active labels counted here are used to give an idea of the share of active versus inactive datapoints, in the next sections the activity calls will be adjusted with the plasma concentrations.

Table 3.5 Size of bioactivity datasets for drugs from the AE datasets, contrasting the set of measured data with the set of measured plus predicted data. Predicted active is defined as the predicted probability of activity > 0.6 for any of the predicted concentrations (100,10,1 and 0.1 μM) for the purpose of this dataset count. A pChEMBL value of 6 corresponds to a concentration of 1 μM.

	Measured bioactivity dataset	**Measured + predicted bioactivity dataset**
Number of unique parent drugs	1,147	1,491
Number of unique targets	1,044	1,044
Matrix cells with data (%)	4.8	26.6
Drug-target pairs with data (n)	56,909	413,840
Datapoints derived from predictions (%)	-	86.2
Datapoints "active" (pchembl_value ≥ 6) (%)	6.7	1.3

3.3.2 Distribution of compiled plasma concentrations

Plasma concentrations were extracted from publications and ChEMBL, and unbound drug concentrations were calculated where possible. Total and unbound concentrations were available

for 625 and 466 drugs respectively. As expected, the unbound concentrations are lower than the total concentrations (Figure 3.1). The median total plasma concentration is 5.8 (pMolar) with a standard deviation of 1.3, whereas the median unbound concentration is 6.6 (pMolar) and the standard deviation is 1.5 (Figure 3.1). Some of the highest values on the distribution are those around pMolar concentrations of 3 and 4, corresponding to plasma concentrations of 1 mM and 100 μM respectively. As described in the Methods, the drugs with multiple plasma concentration measurements of which the standard deviation was greater than 1 log unit were inspected against original data. This check revealed 8 errors in ChEMBL data, mostly related to unit transcription errors, which were corrected, and while it would not be feasible to check every data point when using data from other sources, this illustrates that one of the limitations of using data from previous compilations and databases is potential transcription errors. Another limitation of this data is that different routes of administration are all included but respective plasma concentrations may differ *(36, 176)*. Lastly, the unbound plasma concentrations are based on the f_u and PPB, which are mostly measured *in vitro (36)* and may not reflect *in vivo* values perfectly.

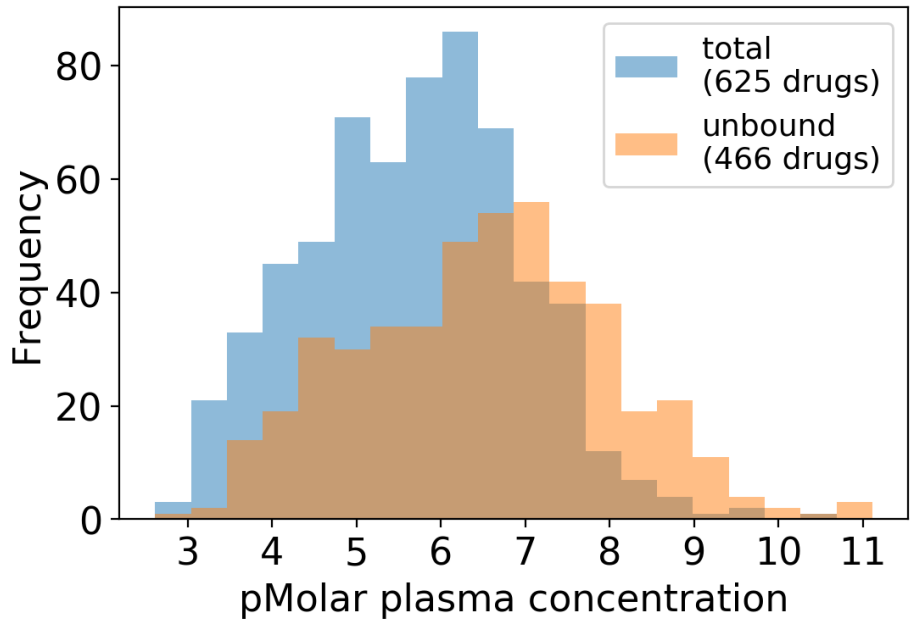

Figure 3.1 Distributions of the total and calculated unbound plasma concentrations available for the drugs from the AE datasets prepared in chapter 2. In case of multiple measurements for the same drug, the median was taken.

Looking at the difference between the total and unbound concentrations, Figure 3.2 shows the distribution of the difference between the total and unbound concentrations per drug. For the

majority of drugs, the difference is less than 1 log unit but there are cases of over 3 log units difference, which seems considerable and would be expected to impact further analysis.

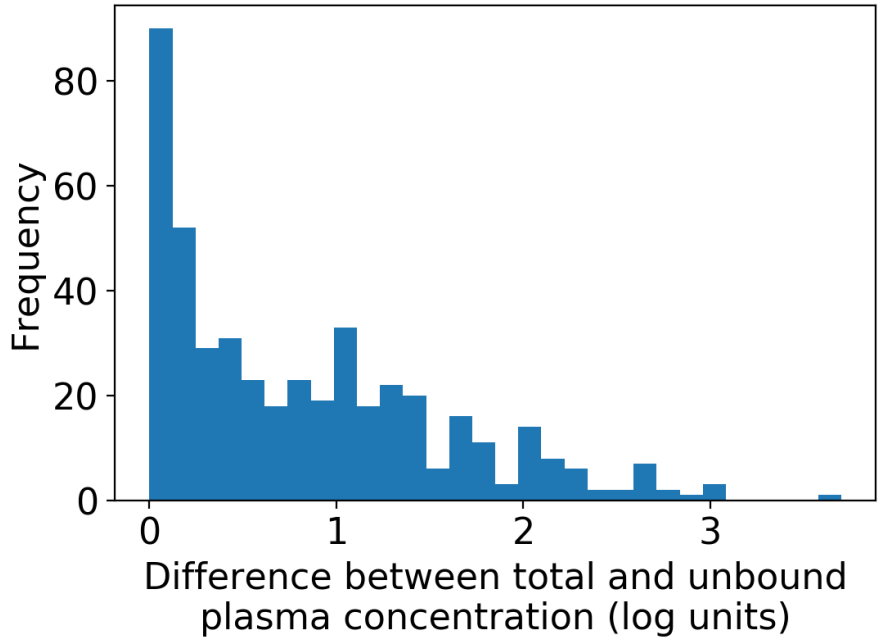

Figure 3.2. Distribution of the difference between the total and unbound plasma concentrations per drug

3.3.3 Integration of bioactivity data and plasma concentrations

3.3.3.1 Effect of using plasma concentrations on the size of the dataset

Plasma concentrations were only available for a subset of all drugs, therefore reducing the size of the dataset for further analysis. To quantify this effect on the size of the dataset, Table 3.6 shows the number of datapoints when using the plasma concentrations as a percentage of the dataset using the constant cut-off. As expected, requiring plasma concentrations in the analysis reduces the size of the dataset. Overall, 63% and 51% of the drug-target pairs are retained when using the total and unbound plasma concentrations respectively. While the number of unique drugs is reduced by up to 7%, the number of unique targets when using the total and unbound concentrations is reduced to 83% and 63% of the dataset respectively. Furthermore, requiring unbound concentrations restricts the dataset more than requiring total concentrations due to limited data availability. Overall, it is concluded that requiring drug plasma concentrations reduces the available drug-target bioactivity datapoints by around 40-50%. The rest of the results in this chapter focus on the activity calls adjusted with the unbound plasma concentrations.

Table 3.6. The percentage of drug-target pairs, unique drugs and unique targets retained when using the drug plasma concentrations compared to using the dataset with pChEMBL value ≥ 6 cut-off. All datasets compared include target predictions.

	Total plasma concentration	Unbound plasma concentration
Drug-target pairs (%)	63	51
Unique drugs (%)	95	93
Unique targets (%)	83	63

3.3.3.2 Effect of using unbound plasma concentrations on the target class distribution

To examine the effect of using unbound plasma concentrations on the overall distribution of bioactivity datapoints across target classes, Figure 3.3 compares the target class distribution of all bioactivity datapoints to those with unbound plasma concentrations available. Although the percentage of kinase datapoints is slightly reduced and the percentage of Family A G protein-coupled receptor datapoints is increased, the distributions are similar (Figure 3.3). Thus, it appears that the subset of the dataset with unbound concentrations available is roughly representative of the complete bioactivity dataset and no large bias has been introduced at this stage.

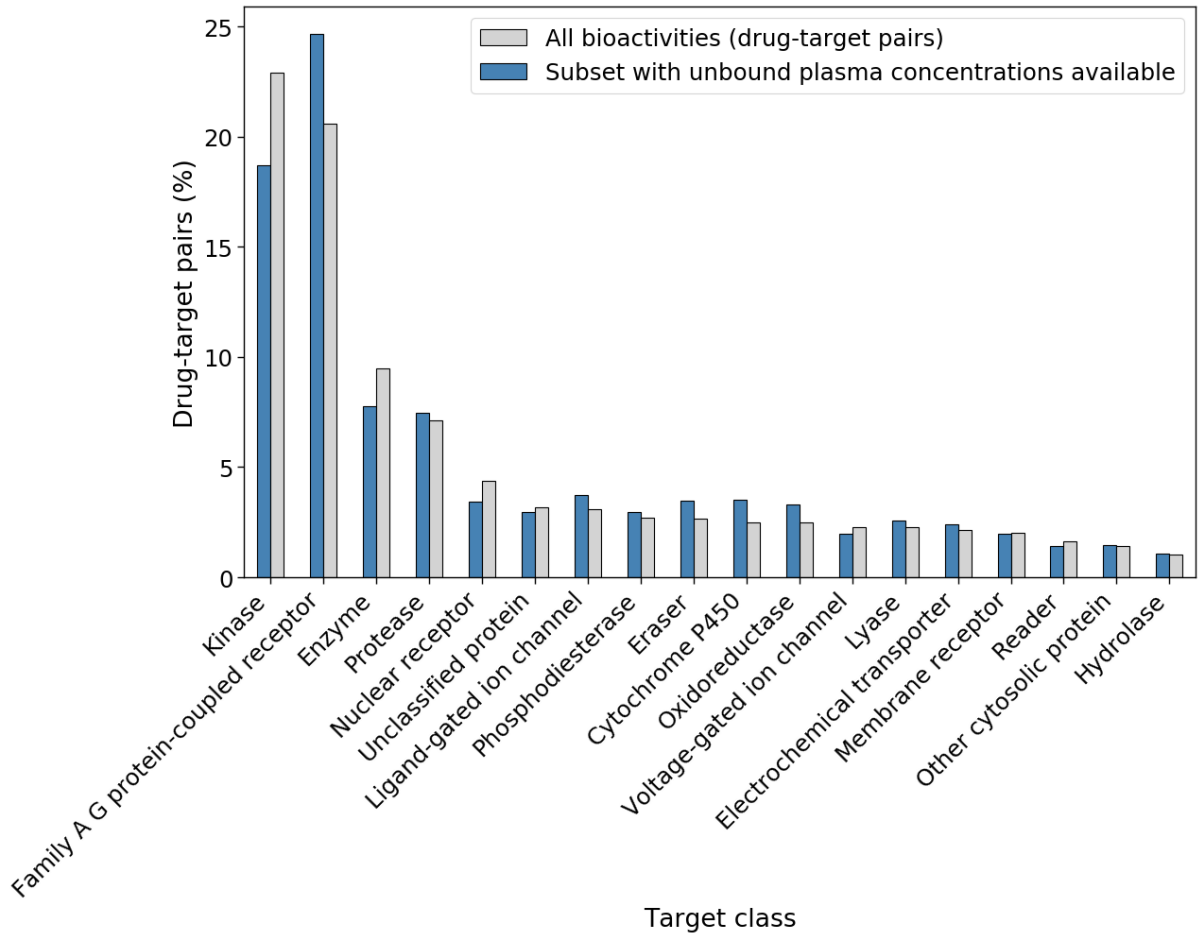

Figure 3.3. Distribution of all bioactivity datapoints across target classes, compared to the datapoints with unbound plasma concentrations available. The distributions are similar, so the subset with unbound plasma concentrations seems to be a representative sample in terms of target classes. The target class shown is the second level of the ChEMBL target hierarchy if available (e.g. lyase), otherwise only the first level is shown (e.g. enzyme). Only target classes containing more than 1% of datapoints are shown.

3.3.3.3 Effect of using unbound plasma concentrations on activity calls

Using the plasma concentrations to derive adjusted activity calls results in 1.0% of all datapoints being classified as active, which is similar to the 1.3% in the dataset based on using the constant bioactivity threshold. Also similar is that the percentage of adjusted activity calls classified as active is higher among the experimental (3.6%) than predicted datapoints (0.1%), upholding the observation that the predicted datapoints primarily add datapoints classified as inactive.

Considering the drug-target pairs that are present among both the baseline and adjusted activity calls, only 0.6% of all datapoints is different, i.e. the activity call is reversed by consideration of the unbound plasma concentration. Of these changes, 68% are in the direction of inactive to active,

resulting in drug-target pairs newly labelled as active, which would not be classified as active when using the constant cut-off. In addition, the changes in the direction of active to inactive affect 19% of active labels among the baseline activity calls. Figure 3.4 contrasts the active drug-target pairs among the baseline and adjusted activity calls, showing the 19% (n=298 drug-target pairs) of active pairs in the constant cut-off dataset are not shared with the active adjusted activity calls. A total of 647 pairs are only assigned active when using the unbound concentrations. This means that 34% (647 out of (1250 + 647)) of active labels among the adjusted activity calls are unique to this scenario, and would be classified as inactive with the constant cut-off. Thus, even though the majority of active labels stays the same after adjustment with the unbound plasma concentrations (1250), the set of active drugs changes considerably when using the unbound plasma concentrations compared to the absolute cut-off. To what extent the adjusted activity calls are more accurate and relevant to *in vivo* AEs will be investigated in the next chapter, which will consider how the targets selected by the different activity calls relate to AEs. At this stage, it can be concluded that the set of active labels is considerably different after integration the with unbound plasma concentrations and this warrants further investigation.

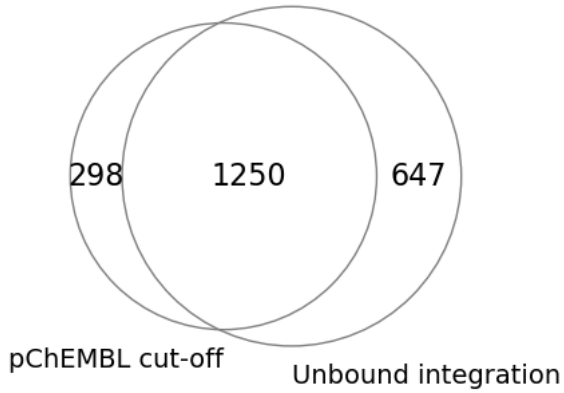

Figure 3.4. Overlap in drug-target pairs classified as active when using the constant cut-off of pChEMBL value ≥ 6 compared to using the ratio of the unbound plasma concentration and the pChEMBL value (adjusted activity calls) as described in the Methods. 19% of the active calls (n=298) when using the constant cut-off are changed to inactive by considering the unbound plasma integration, while a further 647 pairs are changed from inactive to active by the integration, meaning they are unique active calls to the dataset based on using unbound drug plasma concentrations.

3.3.3.4 Effect of using plasma concentrations on activity calls by target class

To investigate how the integration with unbound plasma concentrations affects the activity calls in different target classes, Figure 3.5 shows the distribution of datapoints across target classes. The

top bar for each target class indicates the percentage of all drug-target pairs that fall within that class, whereas the lower bar indicates the percentage of all datapoints changed by the integration that fall within the class (Figure 3.5). For example, bioactivities at kinases comprise around 19% of all datapoints with unbound concentrations available, but only 4% of the activity calls changed by unbound plasma concentrations concern kinase datapoints. Thus, activity calls for kinases are proportionately less affected by plasma concentrations, meaning their baseline and adjusted activity calls are generally the same. The lower bar for each target class in Figure 3.5 furthermore show the direction of change for the datapoints that are affected by plasma concentrations, e.g. for kinase datapoints that are changed, 79% is in the direction inactive to active.

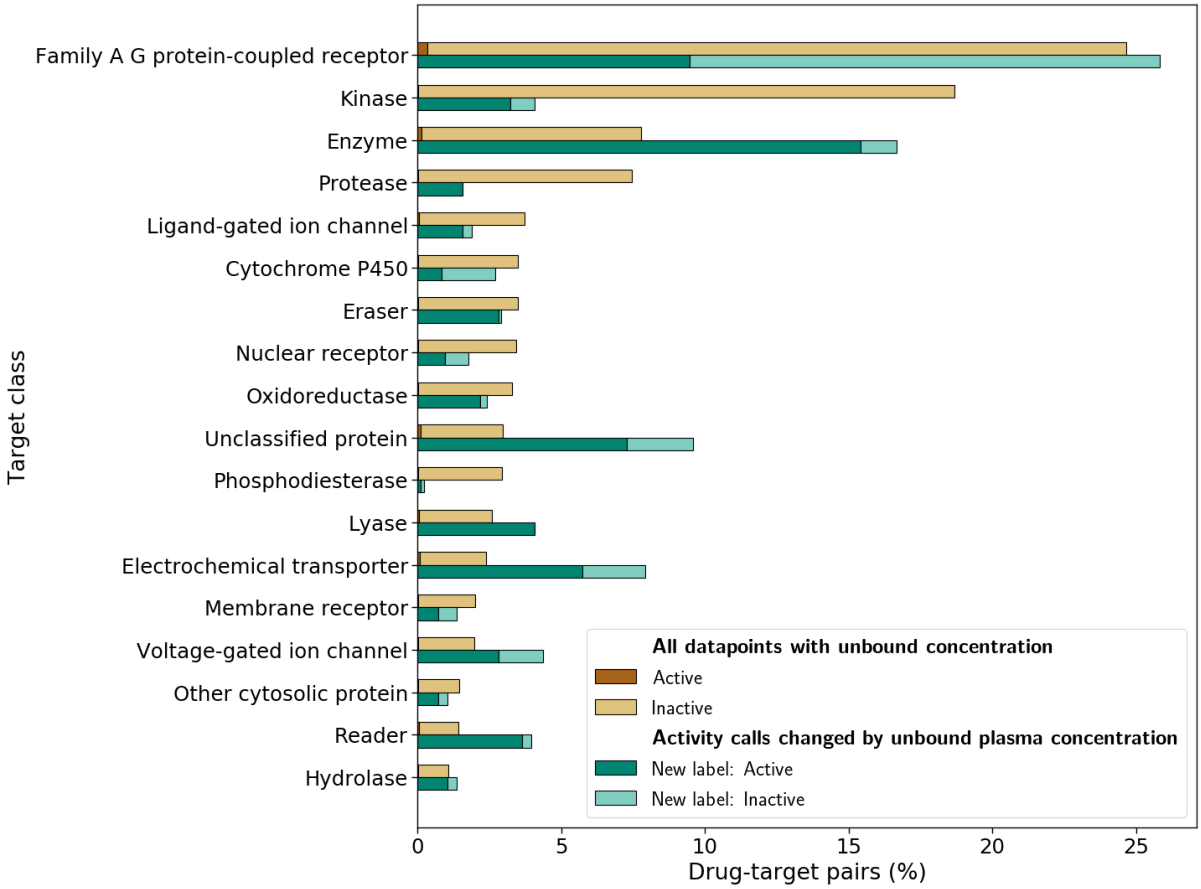

Figure 3.5. Distribution of all bioactivity datapoints by target class for which the unbound drug plasma concentration was available, compared to the activity calls that are changed by the integration of the unbound plasma concentration, with the underlying activity calls shown. The target class is the second level of the ChEMBL target hierarchy if available (e.g. lyase), otherwise only the first level is shown (e.g. enzyme). Activity calls that are changed by the plasma concentrations most often change from inactive to active, except for the family A GPCRs and

cytochrome P450s, for which most of the changes are from active to inactive. Only target classes containing more than 1% of datapoints are shown.

In contrast to kinases, datapoints for enzymes, unclassified proteins, voltage-gated ion channels, lyases, and electrochemical transporters are proportionately more affected by the integration. For example, 17% of datapoints changed by the plasma concentrations are enzyme datapoints, whereas <8% of all datapoints fall within this class (Figure 3.5). Thus, these classes appear more sensitive to the integration with unbound plasma concentrations.

Furthermore, looking at the direction of changes in the lower bar of each class, it can be seen that for the majority of datapoints, adjustments by plasma concentrations occur in the direction from inactive to active (Figure 3.5). In fact, for lyases, proteases and writers, all changes are in the direction from inactive to active (Figure 3.5). Exceptions to this observation are the family A GPCRs and cytochrome P450s, for which the majority (63%) of changes are from active to inactive (Figure 3.5).

In conclusion, the integration with unbound plasma concentrations does not affect activity calls equally in each of the target classes, with enzyme, unclassified protein, and electrochemical transporter datapoints being more frequently changed, whereas kinase datapoints more often stay the same. When datapoints are changed, these are most frequently in the direction from inactive to active, except for family A GPCRs, for which most of the adjustments result in a new inactive label. The next chapter will investigate the relevance of the novel activity calls for relationships to AEs, but if the activity calls using plasma concentrations are more accurate, as would be expected, these results suggest that using a constant bioactivity cut-off results in more false negative activity calls overall, except for family A GPCRs, for which it would result in more false positive calls. This suggests that consideration of unbound plasma concentrations may be able to reduce the rate of false negative activity calls based for most target classes and reduce the rate of false positive activity calls for GPCRs.

3.3.4 Effect of minimum support (≥5 actives) filter

The final step in preparing the drug-target datasets for further analysis is filtering the data for a minimum of 5 actives per target in line with previous studies. However, it was observed that this final step affects the distribution of data across target classes, as shown in Figure 3.6. While the distribution is similar for the SIDER and FAERS datasets, the fraction of kinase datapoints drops

from ~22% in the constant cut-off dataset to less than 5% in the adjusted dataset. Previously, the subset of data with plasma concentrations was shown to only have slightly fewer kinase datapoints (Figure 3.3) than the complete dataset, thus this change is due to the filter for a minimum number of actives, not due to the use of plasma concentrations in general.

To further investigate how the number of active compounds varies by target class, Figure 3.7 inspects the number of active datapoints per unique target across target classes, showing that kinases have a low median number of actives per target, in other words, there are few different compounds available for the same kinase. Looking at the underlying datasets, this is related to the low number of kinases in the DrugMatrix dataset, from which a large part of the measured bioactivity data in this work originates, but which only includes 9 unique kinase targets. The originally downloaded bioactivity data for the drugs from the AE datasets - not restricted to those with plasma concentrations includes 341 unique kinases, but for only 22 out of those there are 5 or more unique drugs measured against them. Contrast this with the Family A GPCRs, for which 106 targets are included in the original bioactivity dataset but of which 59 have at least 5 measured drugs and 48 are included in DrugMatrix.

Thus, the underrepresentation of kinases in DrugMatrix means that the rest of the measured kinase data is derived from scientific literature, which provides measurements for a large variety of distinct kinases (>300) but not sufficient data for different approved drugs measured against the same kinase (Figure 3.7). This shows the value of initiatives such as DrugMatrix which screen a full matrix consistently, compared to literature-derived data. The reason why kinases are most affected may be due to the fact that pXC_{50} datapoints were used for the analysis in this work, which are more rare in ChEMBL for kinases, since kinases are often screened on a large scale using a single concentration, yielding percentage inhibition datapoints instead, e.g. *(180)*.

Due to above reasons, the kinases are disproportionately affected by the minimum support filter of 5 datapoints per target when the available data is halved due to requiring plasma concentrations (Figure 3.7). It can be concluded that being restricted to a smaller dataset of drug-target datapoints with available plasma concentrations results in the underrepresentation of kinases, and therefore associations of AEs with kinases removed from the dataset will not be discoverable in the next stages. One possible way of remedying the loss of kinases would be to include percentage inhibition data from ChEMBL, at least for the kinase class.

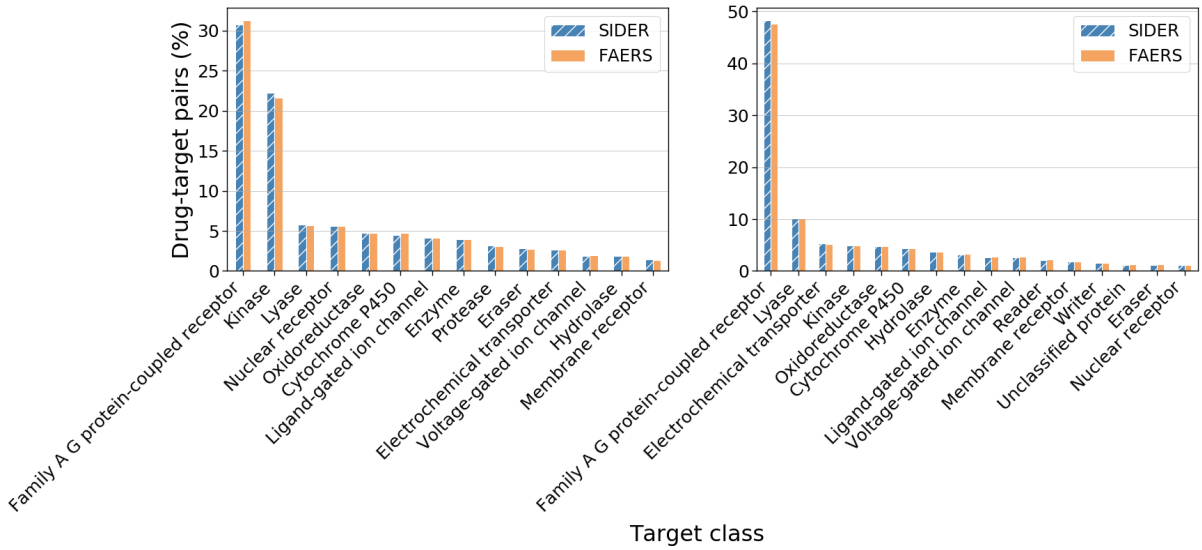

Figure 3.6 Distribution of drug-target datapoints for drugs in the AE datasets across target classes after applying a filter of minimum 5 actives per target on the constant cut-off dataset (left panel) and the dataset using the adjusted activity calls based on unbound plasma concentrations (right panel). The largest change is in the share of kinase datapoints, which is smaller in the latter case.

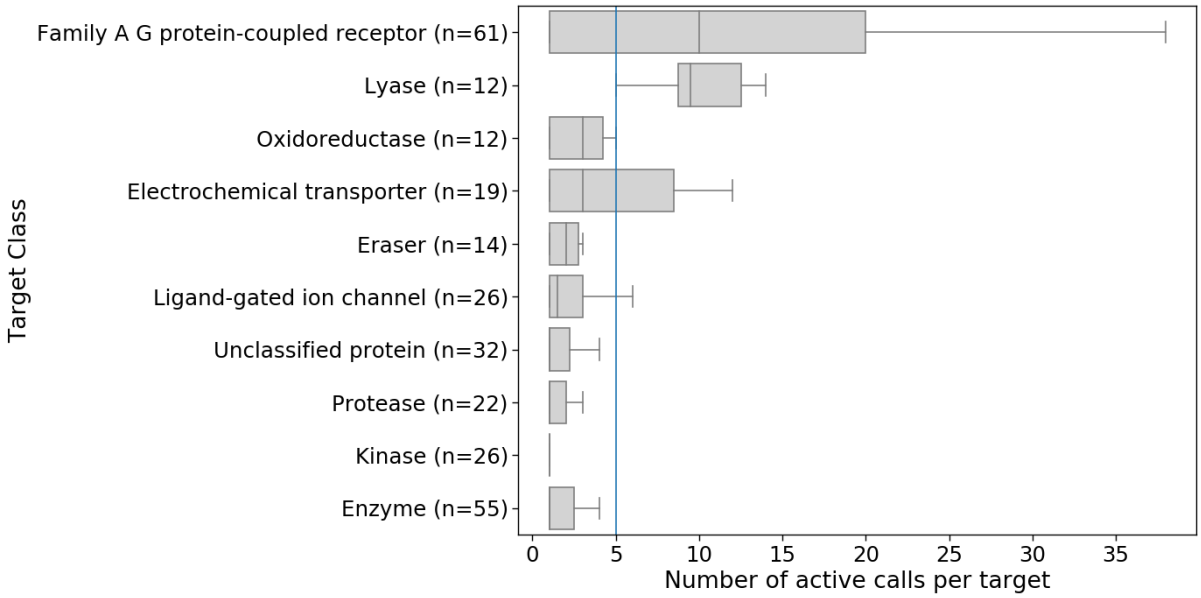

Figure 3.7 Number of active calls per target where activity is defined by the ratio of the unbound plasma concentration and the *in vitro* pXC_{50}. Kinases have a very low median number of actives per target, thus are affected by the minimum support filter of 5 actives per target, highlighted by the vertical line. The number of unique proteins in each target class is shown in parentheses. Boxplots show the median and 1.5 times the IQR.

3.3.5 Number of bioactivity datapoints for further analysis of FAERS and SIDER datasets

For the final datasets that will be taken forward to the next chapter, Table 3.7 shows the number of datapoints in the baseline dataset and the adjusted dataset for FAERS and SIDER. In both cases, the adjusted dataset is around half the size or smaller than the baseline dataset. The number of bioactivity datapoints and unique drugs in the FAERS datasets is higher than that in the SIDER datasets, while the number of targets is the same (Table 3.7). The total number of unique drugs in the final datasets, now combining FAERS and SIDER, for the constant cut-off and adjusted cut-off scenarios are reported in Table 3.8, showing that drugs are included based on the presence of numeric measurements and plasma concentrations, inactive 'activity_comments', or consistent negative predictions. In conclusion, the dataset using unbound plasma concentrations is around half the size of the constant cut-off dataset in terms of drug-target pairs and the number of drugs with bioactivity data is slightly higher for FAERS than SIDER.

Table 3.7. Number of datapoints in the baseline bioactivity dataset compared to the adjusted dataset based on the integration of *in vitro* bioactivities and unbound plasma concentrations, both of which include predictions, split by the two AE datasets FAERS and SIDER.

	pChEMBL cut-off (pChEMBL value ≥ 6)		**Unbound plasma concentrations**	
	FAERS	SIDER	FAERS	SIDER
Drug-target pairs (n)	85,155	70,953	34,247	27,063
Unique drugs (n)	1,173	1,001	1,019	873
Unique targets	234	234	104	104

Table 3.8 Number of unique drugs in the datasets, showing they are included based on numeric data, based on inactive 'activity_comments' in ChEMBL or based on inactive predictions

	pChEMBL cut-off (n drugs)	**Unbound plasma concentrations (n drugs)**
Numeric bioactivity measurement	1044	Integrated with plasma concentration: 466

Included based on 'inactive' activity_comment only	98	448
Included based on consistent negative prediction at each PIDGINv3 threshold	344	357
Total unique drugs	**1,486**	**1,271**

3.4 Discussion and conclusions

The main contribution of this chapter is the assessment of the impact of integrating unbound drug plasma concentrations with bioactivities on resulting bioactivity calls, which has not previously been reported on a large scale. It was found that requiring unbound plasma concentrations to derive activity calls reduces the size of the drug-target dataset by about 50%. Furthermore, most of the adjustments resulting from consideration of the plasma concentrations occur in the direction inactive to active, thus creating novel active labels, except for the family A GPCRs, for which most changes are from active to inactive. The comparison of active labels among the baseline and adjusted activity calls shows that around 20% of the active calls made using the constant cut-off are changed by consideration of the unbound plasma concentration, and 34% of active-labelled adjusted activity calls are unique to this scenario. At this stage, it can be concluded that the number of changes to active labels as a result of integrating unbound plasma concentrations is not negligible. The next more important question is to what extent this matters for selecting targets relevant to AEs, and this will be investigated in the next chapter. The distribution of datapoints across target classes provides a background against which to interpret these further results, such as the expected absence of AE associations with kinases due to a low number of datapoints, and the prominence of GPCRs, being the largest class of bioactivity datapoints.

Based on the findings in this chapter, it can be concluded that target predictions make a large contribution to the overall bioactivity dataset, comprising around 85% of all datapoints and increasing the density of the drug-target matrix to over 26% instead of 5% with measured datapoints only. It is inevitable that there are more uncertainties in predicted than experimental datapoints, and some of the predictions may be incorrect. This was mitigated as much as possible

by choosing relatively high thresholds for the target prediction applicability domain and model performance in cross-validations.

Limitations of the bioactivity data used include that it is sparse and incomplete *(107)*, and lacks complete information on functional effects *(19, 107)*, which therefore were not taken into account in the current work. Limitations of the drug plasma concentration data derived from literature include that they are a mixture of 'normal' therapeutic concentrations, some of which are actually minimal effective concentrations *(176)*, and C_{max}. Furthermore, the plasma concentrations are not patient-specific because plasma concentrations are not generally included on reports in FAERS. However, it is known that there is inter-individual variation in plasma concentrations *(14, 36, 176)*, which may play a role in ADRs but could not be taken into account in this work. In addition, only one plasma concentration was used per drug, not distinguishing between different drug indications for the same drug that may require different doses. Moreover, the focus on therapeutic drug plasma levels excludes effects due to higher concentrations as a result of drug interactions or overdose. Overall, the aim of using plasma concentrations is to refine the activity calls based on a constant cut-off for *in vitro* potency, acknowledging above limitations of the pharmacokinetic data available for computational analysis.

4 Identification of protein targets associated with adverse events

4.1 Introduction

This chapter will build on the drug-AE datasets prepared in chapter 2 and the drug-target dataset prepared in chapter 3 to compute target-AE relationships, which is the ultimate aim of this work. The workflow from the previous chapters and how these feed into the current chapter is summarised in

Figure 4.1.

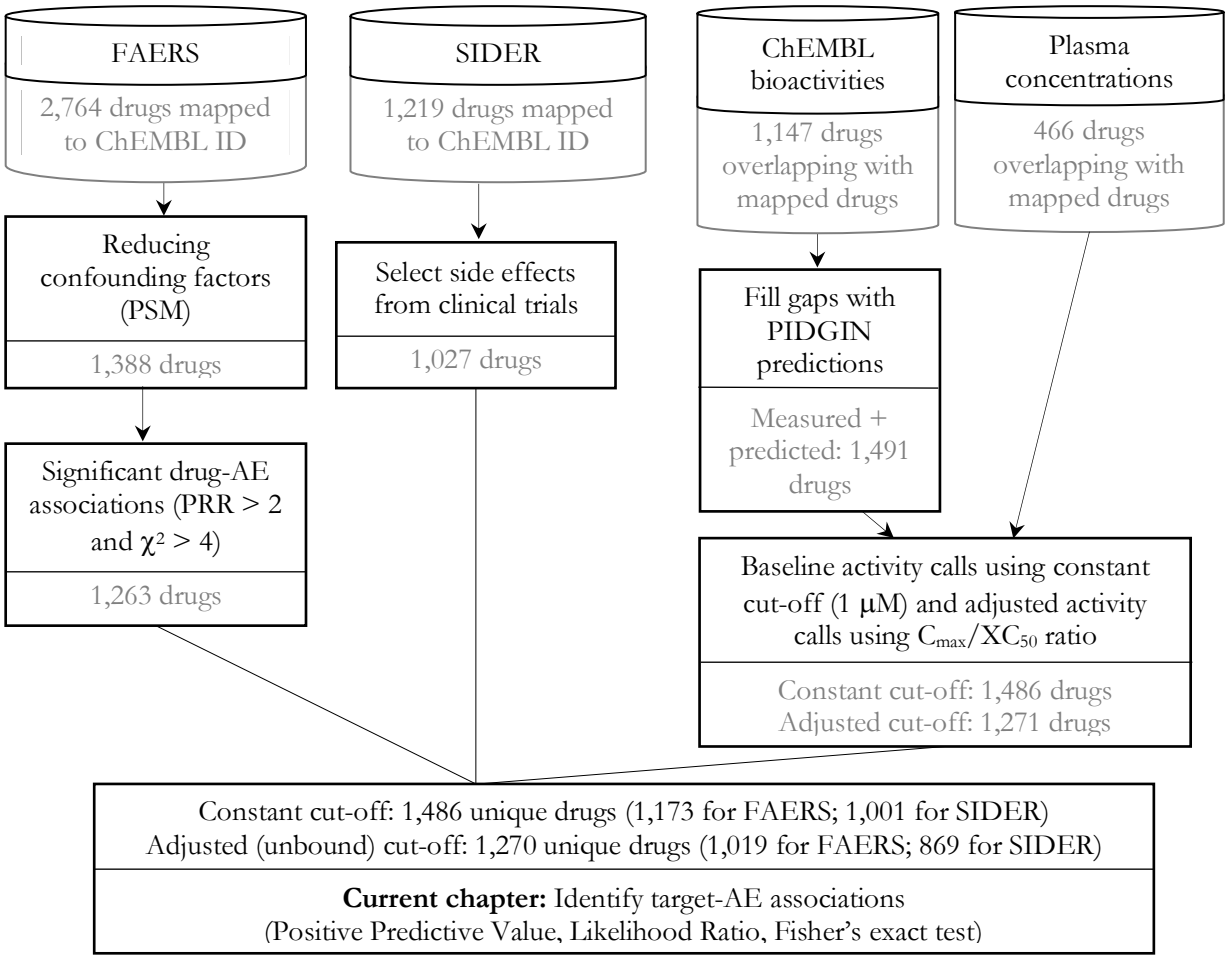

Figure 4.1 Workflow summary. The work from the previous chapters now forms the basis of this chapter with the final aim of identifying target-AE associations.

Deriving target-AE relationships may help to identify targets that are responsible for causing AEs and therefore guide the design of future safety pharmacology screens to anticipate AEs in drug discovery. There is a lack of publicly available systematic annotations of target-AE relationships, which will be addressed in this chapter by using the statistical metrics introduced in the Introduction to identify and quantify overrepresented target-AE combinations. Knowledge of *quantified* associations, such as predictive values, may be able to help the interpretation of secondary pharmacology screening results in terms of the likelihood of *in vivo* effects upon observing *in vitro* activities.

While previous studies have considered drug plasma concentrations alongside secondary pharmacology results in case studies of a handful of target-AE combinations, no study to date has systematically compared the impact of using drug plasma concentrations on derived target-AE relationships, thus this chapter will begin by examining this. Next, the distribution of association metrics and predictive values across all target-AE relationships is presented, and explored across target classes and SOCs. This provides a large-scale overview of target-AE associations.

Then, the strongest individual associations will be considered in the context of previous literature, before comparing the target-AE associations derived in this work to those specifically reported on published safety pharmacology panels. However, focusing on individual target-AE associations has the limitation of not considering cases in which an AE may be caused independently by multiple targets. Thus, this chapter will also explore the value of considering combinations of targets associated with the same AE.

The final analyses in this chapter will assess the distribution of individual derived target-AE associations across SOCs in order to reveal which types of AEs may be predictable from secondary pharmacology screens, and which types of AEs are rarely related to *in vitro* bioactivities, highlighting potential areas with more apparent versus limited translation between *in vitro* and *in vivo* effects. Lastly, the target-AE associations based on SIDER and FAERS will be compared, thereby for the first time assessing their differences with respect to target-AE associations.

4.2 Materials and methods

4.2.1 Identification of significantly associated target-AE combinations

Building on the datasets prepared in chapters 2 and 3, for each target, the drugs from the AE datasets with available measured or predicted bioactivity data were identified and counted in a contingency table as illustrated in Table 4.1. Association with the AE is based on the PRR for FAERS or the presence of the AE in SIDER, as described in chapter 2. Activity at the target is based on either the constant bioactivity cut-off (baseline activity calls) or the C_{max}/XC_{50} ratio (adjusted activity calls), as described in chapter 3 (Table 4.1). Only measured and predicted bioactivity data is used, and inactivity at targets is explicitly not assumed, resulting in a different total number of drugs counted in the contingency table for each drug-target combination.

Table 4.1 Contingency table for the association between measured and predicted bioactivities of drugs and their associations to AEs. Each cell contains the count of the number of drugs in that category.

		Activity at target (Using C_{max}/XC_{50} ratio or a constant cut-off of $IC_{50} < 1$ μM)	
		Active	**Inactive**
Association to AE	**Associated** (PRR > 2 and $\chi^2 > 4$)	True positives (TP)	False negatives (FN)
	Not associated	False positives (FP)	True negatives (TN)

The Scikit-learn version 0.21.3 confusion_matrix function *(162, 181)* was used to obtain the number of true negatives (TN), false positives (FP), false negatives (FN) and true positives (TP). The Fisher's exact p-value was calculated using the SciPy stats version 1.3.2 fisher_exact function *(163)*.

The LR+ was calculated as LR=sensitivity/(1 – specificity) *(141, 145)*. For all pairs with a LR > 1, indicating a positive association between the AE and activity at the target, the Fisher's exact p-values were corrected for multiple testing with the Benjamini-Hochberg method using the Statsmodels 0.9.0 multipletests function *(182)*. Associations were considered statistically significant if they have a corrected p-value ≤ 0.05.

The PPV was obtained by dividing the number of drugs that were active *and* associated with the AE by the total number of drugs active at the target *(141, 143)*.

The value-added PPV was calculated by subtracting the prevalence, i.e. the fraction of drugs measured at a given target that is also associated with a given AE, from the PPV *(141)*.

4.2.2 Compilation of previously reported safety targets and their associated adverse events

The target-AE associations reported for three published secondary pharmacology panels *(19, 20, 62)*, which were presented in Table 1.4 in the Introduction, were compiled by manually extracting protein names and adverse effect descriptions. The protein names were manually mapped to Uniprot protein identifiers in ChEMBL based on name. Arrows and terms used in the original AE description, such as 'enhances', 'induces', 'facilitates', 'exacerbates' etc. were reformatted to 'increased' and terms such as 'inhibits', 'reduces', 'impairs' etc. to 'decreased'. The terms were then submitted to the National Center for Biomedical Ontology Bioportal, specifically the 'Annotate' functionality, for annotation with MedDRA terms *(183)*. All results were manually inspected and selected. Additionally, mappings for terms that did not yield any results in this way were identified by manually querying the terms in the MedDRA web-based browser and selecting appropriate PTs or HLTs *(25)*. The HLTs and SOCs from MedDRA versions 21.1 and 22.1 were obtained using the Hierarchy Analysis function in the MedDRA web-based browser *(25)*. To calculate the recall, associations from FAERS and SIDER in the current study were considered previously reported if the HLT matched the HLT of a previously reported association for the same target, considering the "Primary SOC" obtained from the MedDRA web-based browser *(25)*.

4.2.3 Assessing combinations of targets

To assess combinations of targets in cases where multiple targets were each individually associated with a given AE, all possible sets of the associated targets were listed. Next, drugs that were

measured at all these targets were identified, if any. Using this set of drugs and the logical 'or' to consider bioactivity at any of the constituent targets, the PPV and fraction of AE-associated drugs identified was calculated for the combination overall.

4.3 Results and discussion

4.3.1 Impact of using plasma concentrations on target-AE associations

To assess the impact of using plasma concentrations compared to using a constant bioactivity cut-off on target-AE associations, the number of target-AE pairs that could be studied in either case is first compared. After overlapping the bioactivity data prepared in chapter 3 with the AE data from chapter 2, the number of target-AE pairs that could be considered is shown in Table 4.2. The FAERS dataset contains about four times as many target-AE pairs as the SIDER dataset, both when using a constant bioactivity cut-off and when using the adjusted activity calls. This is consistent with the higher number of unique PTs present in the drug-AE dataset based on FAERS, which was observed in chapter 2. Since there are more unique PTs, this results in a greater number of target-AE combinations, given that the number of targets in the bioactivity datasets was the same for FAERS and SIDER (chapter 3).

Similar to the reductions in available bioactivity data for analysis when using unbound plasma concentrations described in chapter 3, the dataset using the adjusted activity calls also contains around 40-50% fewer target-AE pairs compared to using the baseline activity calls (Table 4.2). Thus, using unbound plasma concentrations limits the number of target-AE combinations that can be evaluated in the study.

Table 4.2. The number of target-AE pairs that were considered in the study

	pChEMBL cut-off (baseline activity calls)		**Unbound plasma concentrations (adjusted activity calls)**	
	FAERS	SIDER	FAERS	SIDER
Target-AE pairs considered (n target-AE combinations)	313,661	89,105	197,236	42,652
Unique targets (n)	182	167	100	79

Unique adverse events (n MedDRA PTs)	3340	1119	3278	982

Of the numbers in Table 4.2, a total of 1,092 (FAERS) and 2,155 (SIDER) target-AE associations were statistically significant based on the constant cut-off and 224 (FAERS) and 315 (SIDER) in the case of the unbound plasma concentrations. Thus, even though more combinations were considered in the FAERS dataset, SIDER yields more significant target-AE combinations, which may be related to the noisier nature of FAERS, as well as to the high number of distinct AEs in FAERS, for which there may not be sufficient reports to result in statistical significance.

Comparing the significant target-AE associations shows that 21% (SIDER) and 66% (FAERS) of target-AE associations found when using unbound concentrations are not significant when using the constant cut-off. Thus, the significant target-AE associations found when using plasma concentrations are not a direct subset of the dataset using a constant cut-off, showing that using plasma concentrations results in a different set of significant target-AE associations.

With regards to the accuracy of the associations, the lack of a gold standard for target-AE associations makes it difficult to evaluate both outcomes *(70)*. Therefore the target-AE associations from the two datasets were evaluated by comparing them to previously reported associations from three reported safety target panels *(19, 20, 62)*. The overall retrieval of these previously reported associations as statistically significant in this study, at the level of HLTs, is 6% using the unbound concentrations and 12% when using the constant cut-off. Thus, using the unbound concentrations results in a lower recall of previously reported associations. This could be related to the fact that plasma concentrations were only available for a subset of the data, resulting in fewer datapoints for analysis (Table 4.2), since this likely reduced the power of this study.

The low recall generally, being only 12% when using the constant cut-off, could be related to differences in the way AEs are recorded, for example, the review by Bowes et al. focuses on biological effects such as "increase in blood pressure" *(19)*. In patients, such effects may not be immediately noticed but could ultimately result in a stroke, which could end up being the reported AE in pharmacovigilance activities. As a result, the exact AE terms, even at the HLT, would not match despite the effects being biologically related. This would result in lack of recall. Similarly, the effects described in Lynch et al. *(62)* focus on pathological effects from animal studies, of which some, such as organ histopathology, are unlikely to be observed in humans. Alternatively,

effects may only be observed at high doses that may be used in animal studies but are not used clinically, further contributing to low recall of previously reported effects in this work.

However, inspecting the distribution of the measures of association for those target-AE pairs that are retrieved shows that those from the dataset based on unbound plasma concentrations have larger LRs, indicating that adjusted *in vitro* target activities are more useful for identifying AEs than bioactivities based on the constant cut-off (Figure 4.2). Similarly, the PPVs and value-added PPVs of the associations that are retrieved are higher when using unbound plasma concentrations (Figure 4.2). This indicates that using plasma concentration results in fewer false positive signals compared to using the constant cut-off, and is overall more precise. However, the fraction of drugs associated with the AE per target, which would be the 'detection' rate of AE-associated drugs if the target is used as a predictor of AEs, is lower for the target-AE associations based on the adjusted activity calls (Figure 4.2). This indicates that the higher LRs and greater precision when using plasma concentrations comes at the cost of lower recall.

One interpretation of these results is that using plasma concentrations retrieves known signals more strongly and precisely, but that recall could be limited by the amount of data available. Despite the lower overall recall, some familiar associations, such as hERG-torsade de pointes and 5-HT_{2B}-cardiotoxicity in the FAERS dataset, as well as dopamine D2-galactorrhoea in both the FAERS and SIDER datasets, were only significant when using plasma concentrations and not when using the constant cut-off. Since the plasma concentrations are a novel aspect of the current study and the results show that they retrieve known signals with greater LRs and PPVs, the remainder of this chapter will focus on the results obtained with adjusted activity calls.

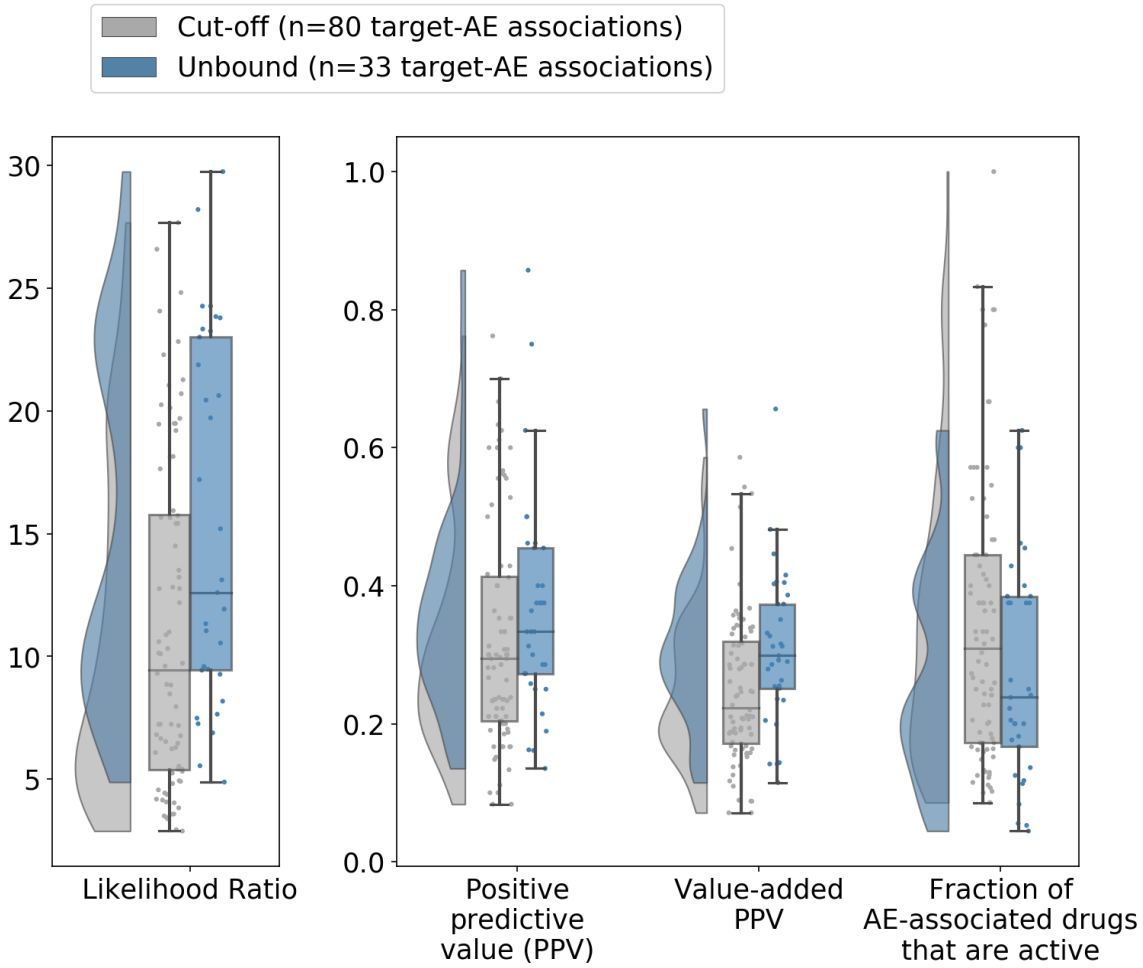

Figure 4.2 Quantification of target-AE associations reported in previous studies that were retrieved as significant in the current work when using a constant bioactivity threshold (pChEMBL ≥ 6) to define active drug-target pairs, versus using the ratio of the unbound plasma concentrations over the *in vitro* bioactivity to do so. Using the unbound plasma concentrations retrieves known associations with greater strength of association (median LR) and the associations are more precise (median PPV). However, this comes at the cost of a lower recall, as only 33 versus 80 previously reported target-AE associations are retrieved, and a lower detection rate of AE-associated drugs (fraction of AE-associated drugs that are active). Boxplots show the IQR including the median and whiskers extend to 1.5 times the IQR.

4.3.2 Quantification of associations between the activity of drugs on proteins and AE reporting

Activity calls adjusted using the unbound plasma concentrations were evaluated as predictors of AEs by analysing to what extent statistically significant target-AE associations provide information about the AE. The median LR for target-AE associations from FAERs is 11.8 and for those from

SIDER the median LR is 16.4 (Figure 4.3). LRs higher than 1 indicate that *in vitro* activities at the target are observed more frequently among AE-associated drugs than other drugs, thus the median LRs being above 10 indicate that being active at these increases the probability of the *in vivo* AE (Figure 4.3) *(147)*. Next, the results focus on the PPV, the fraction of active drugs that are associated with the AE, because this relates the *presence* of bioactivity to the *presence* AEs. The median PPV for significant target-AE associations from SIDER is 0.38 with a standard deviation of 0.2, while FAERS associations have a median PPV of 0.23 (standard deviation=0.1), meaning that across significant target-AE associations ~23-38% of drugs active at the target are associated with an *in vivo* AE (Figure 4.3). The observation that PPVs are lower for FAERS compared to SIDER means that bioactivities have higher false positive rates for AEs in FAERS. Thus, although the fraction of drug associated with the AE, i.e. the detection rate of AE-associated drugs, or sensitivity, is higher for FAERS, this is at the cost of lower precision (PPVs). The lower LRs and PPVs for FAERS compared to SIDER could be due to FAERS being noisier as a result of the way the way AEs are reported, making it more difficult to identify target-AE associations in the dataset.

The value-added PPV indicates the additional information provided by the protein target activity over the prevalence to predict AEs. This is because higher PPVs are generally seen with higher prevalence, referring in this study to the fraction of drugs with bioactivity data that are associated with a certain AE *(184)*. The value-added PPVs follow a similar distribution to the PPVs, except for associations with high PPVs, such as 1.0, which are reduced. (Figure 4.3). This shows that the bioactivities provide additional information and that the PPVs are not primarily driven by the prevalence itself. The prevalence (fraction of drugs associated with AE) is plotted for reference, showing that for statistically significant associations the prevalence is generally less than 5% (Figure 4.3).

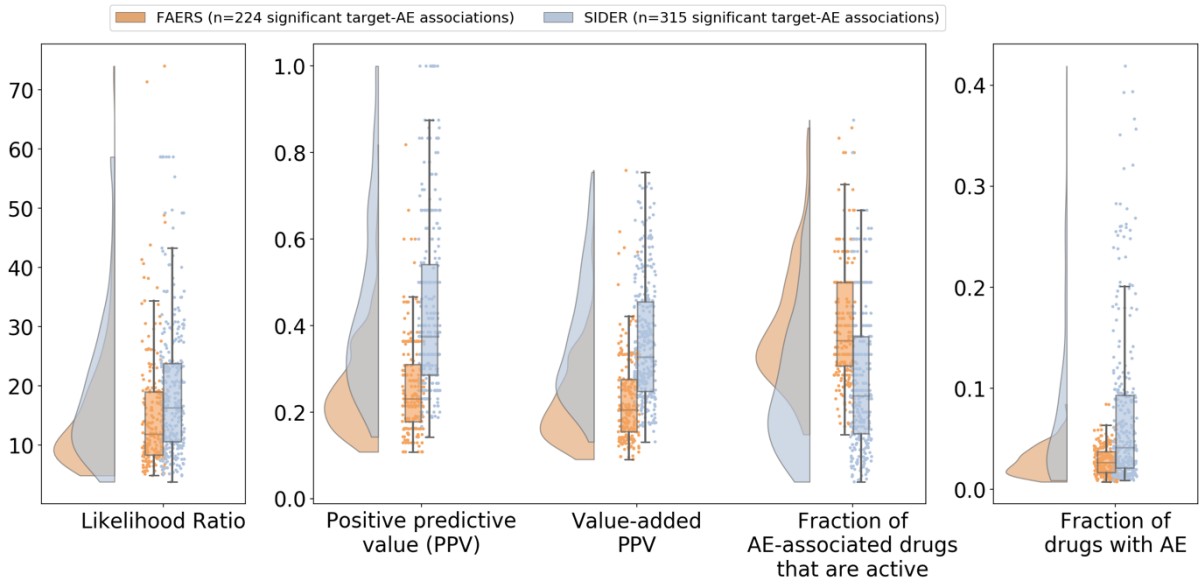

Figure 4.3 Quantification of statistically significant target-AE associations using the LR, PPV and value-added PPV. Associations in SIDER have higher LRs, with a median of 16.4 compared to 11.8 for FAERS. Similarly, the PPVs are higher in SIDER, with a median of 0.38 compared to 0.23 for FAERS. Boxplots show the IQR including the median and whiskers extend to 1.5 times the IQR.

4.3.2.1 Positive Predictive Values by target class

To investigate how PPVs differ by target class, Figure 4.4 shows the distribution of PPVs of significant target-AE associations by target class, now combining the FAERS and SIDER data. The highest PPVs (up to 1.0) are observed in the lyase and family A GPCRs, which are the target classes containing the highest number of associations. Oxidoreductases and electrochemical transporters generally have lower PPVs (up to ~0.6-0.8), and there are also fewer target-AEs associations in these classes (Figure 4.4). In addition, a small number of membrane receptors – those not further classified in the ChEMBL hierarchy – also have lower PPVs (up to 0.38) (Figure 4.4). The target-AE pairs that were reported on published safety target panels *(19, 20, 62)* and retrieved in the current study are also shown in Figure 4.4, showing that some previously reported associations have low PPVs, such as the relationship between the muscarinic M3 receptor (a family A GPCR) and miosis, which has a PPV of 0.25 (Figure 4.4). This shows that some target-AE associations that have been previously reported were retrieved as significant in this study but have low PPVs, possibly due to AEs being known but not often reported. In other cases, the mapping categories used for annotating associations as previously reported, i.e. the MedDRA HLTs, are very broad such as 'neurological signs and symptoms' which results in precision being lost. At the

same time, the results show that the current study identifies many novel target-AE associations which might include mechanistic target-AE links.

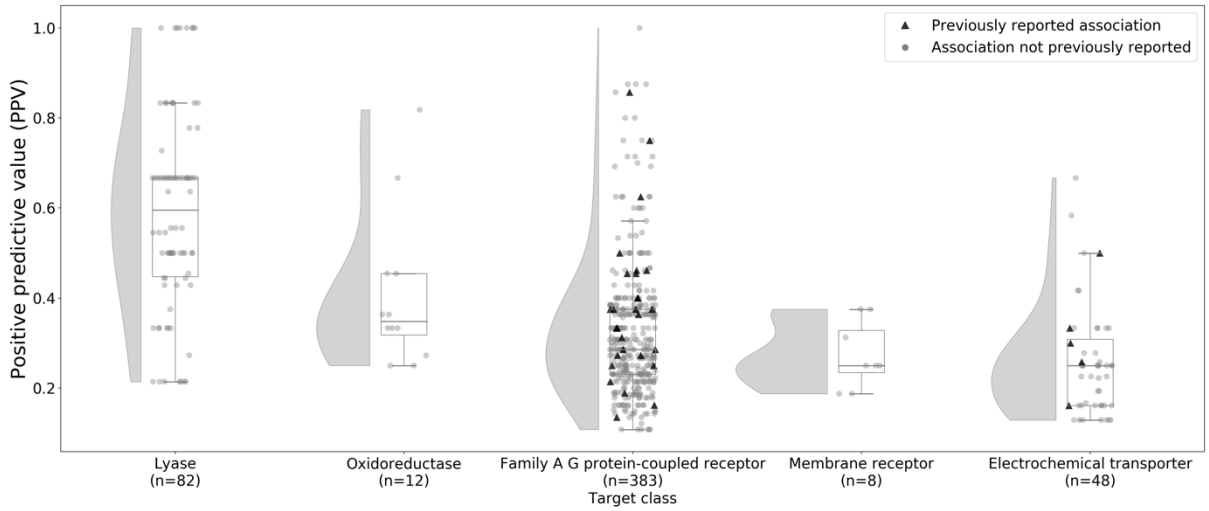

Figure 4.4 PPVs of significant target-AE associations by protein class of the target, for classes with more than 4 target-AE associations. Classes correspond to the second level of the ChEMBL target hierarchy except 'membrane receptors', which is the highest level since those targets are not further classified. The highest PPVs occur in the lyase and family A GPCR classes, which also contain the highest number of target-AE associations. Boxplots show the IQR including the median and whiskers extend to 1.5 times the IQR.

4.3.2.2 Positive Predictive Values by System Organ Class (SOC)

The distribution of PPVs for target-AE associations across MedDRA SOCs of the AE is examined in Figure 4.5. The range of PPVs within most SOCs is wide, with the highest median PPV of 0.56 occurring for 'blood and lymphatic system disorders', and the lowest median PPV of 0.21 for 'musculoskeletal and connective tissue disorders' (Figure 4.5). PPVs above 0.8 only occur in the large SOCs such as 'nervous system disorders' and 'gastrointestinal disorders', which are also the SOCs with the highest number of AEs in the underlying datasets (Figure 4.5). Otherwise, the PPV does not follow an easily interpretable pattern by SOC (Figure 4.5).

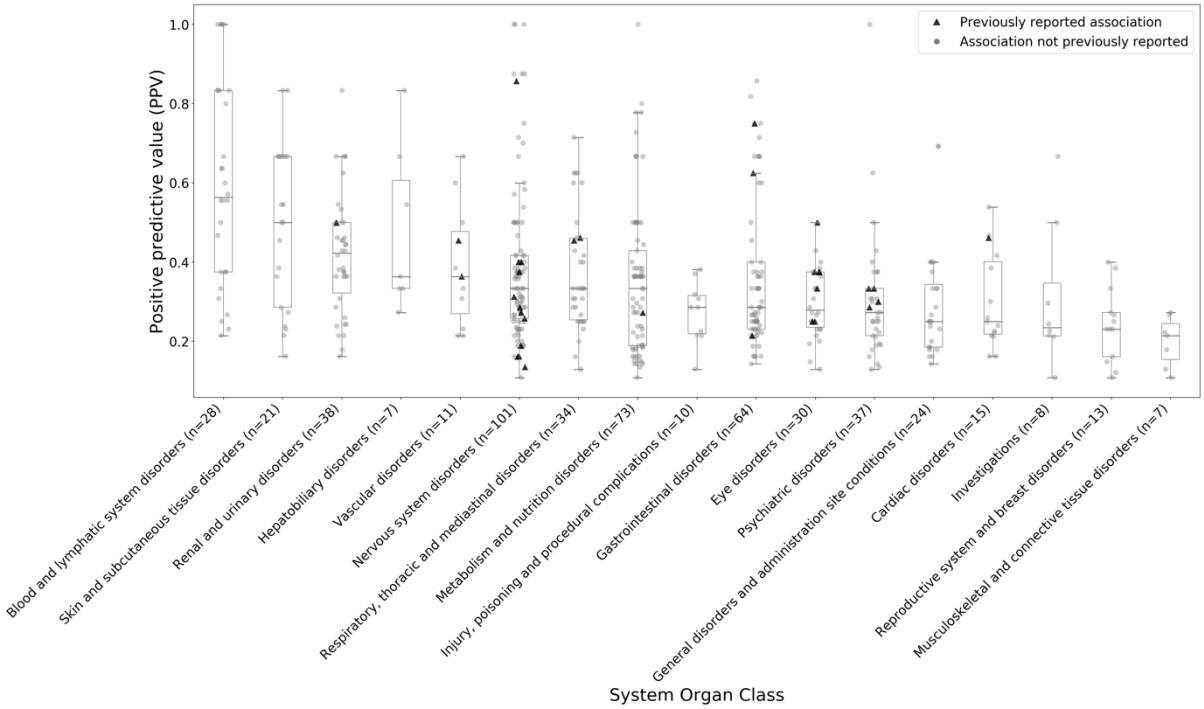

Figure 4.5 PPVs of target-AE associations across MedDRA SOC of the AE. The highest median PPV of 0.56 occurs in 'blood and lymphatic system disorders'. PPVs above 0.8 only occur in the SOCs with the most target-AE associations, such as nervous system and gastrointestinal disorders. Boxplots show the IQR including the median and whiskers extend to 1.5 times the IQR.

4.3.3 Trade-off between Positive Predictive Values and detection of AE-associated drugs

The relationship between the PPV and the fraction of AE-associated drugs that are active at the target, which is the sensitivity or the fraction of AE-associated drugs that would be detected by bioactivity at the target if used a single predictor of the AE, is analysed in Figure 4.6, showing a clear inverse relationship between the two variables. For example, the PPV of the muscarinic M2 receptor activity for tremor is 0.86, but the fraction of AE-associated drugs of around 0.05 indicates that only 5% of drugs associated with this AE in the dataset are active at the receptor (Figure 4.6). Conversely, activity at dopamine D2 receptor would detect 57% of drugs associated with hyperprolactinaemia, but has a false positive rate of 86% (PPV=0.14) (Figure 4.6). Specifically, there are few target-AE associations with a fraction of AE-associated drugs that are active above 0.5, and at the same time high PPVs (Figure 4.6), which would correspond to a simultaneous high sensitivity and a low false positive rate, and be most useful in practice. Thus, no single bioactivity serves as a strong indicator of clinical and post-marketing AEs in the dataset studied. There are many possible reasons for this – e.g. a low fraction of AE-associated drugs would be detected per target if multiple mechanisms involving different targets lead to the same

AE, which will be explored further in section 4.3.7 below. On the other hand, low PPVs can result from pharmacokinetic behaviour of drugs such as lack of blood-brain barrier crossing, leading to certain AEs not being observed in practice.

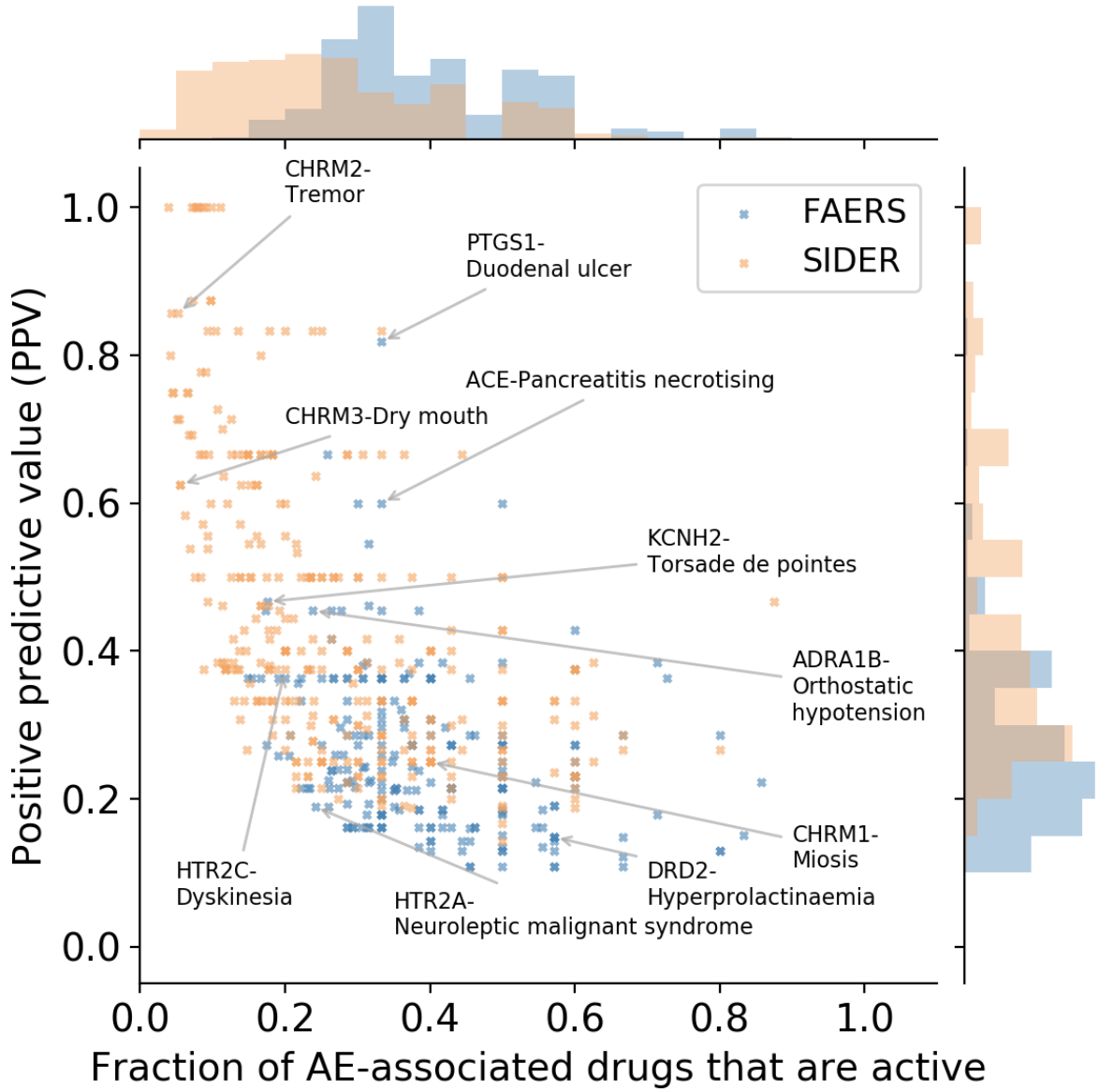

Figure 4.6 Trade-off between the PPV of significant target-AE associations and the fraction of AE-associated drugs that are active at the target and would therefore be 'detected' by the target bioactivity. Activity here is based on the integration of the unbound plasma concentration with measured and predicted *in vitro* bioactivities. Target-AE pairs with high PPVs tend to have low fractions of AE-associated drugs that are active, meaning only a small share of all drugs associated with the AE would be detected by bioactivity at the target. Alternatively, associations with high fractions of AE-associated drugs that are active tend to have low PPVs, indicating a high false positive rate for that target-AE pair. CHRM2: Muscarinic acetylcholine receptor M2,

PTGS1: Cyclooxygenase-1, CHRM3: Muscarinic acetylcholine receptor M3, ACE: Angiotensin-converting enzyme, KCNH2: hERG, ADRA1B: α-1b adrenergic receptor, CHRM1: Muscarinic acetylcholine receptor M1, DRD2: Dopamine D2 receptor, HTR2A: Serotonin 2a (5-HT2a) receptor, HTR2C: Serotonin 2c (5-HT2c) receptor.

One of the limitations of this analysis that may have resulted in the low PPVs, low fractions of AE-associated drugs detected and low recall of previously reported associations is that the analysis is primarily based on marketed drugs. These drugs have already undergone safety screening, so the termination of problematic candidates early on has biased the data available in such a way that strong associations with safety targets, e.g. adverse effects observed in animal studies, may not be apparent when studying patient data. Another potential reason for missing previously reported effects is the incomplete consideration of dose in the current work due to consideration of the therapeutic plasma concentrations only. Effects previously reported in e.g. Bowes et al. *(19)* generally do not list the dose at which effects are expected. This may explain some of the discrepancies between the results presented here and previous studies. Another limitation resulting from the work being based on data from marketed small molecules for a wide range of indications, is that the PPVs do not necessarily apply to other types of drugs such as antibodies or drugs in a specific therapeutic class, because the prevalence of AEs may be different.

Few studies have directly reported PPVs of secondary pharmacology targets to compare these results to. A previous study by Pollard et al., studying marketed drugs, found a PPV of 1.0 for the relationship between hERG and QT interval prolongation when the margin between the *in vitro* hERG IC_{50} and C_{max} was less than 10-fold, which is similar to the threshold in the current work *(71)*. The difference between this PPV and the one in the current study (PPV=0.47 for TdP) is explained by QT interval prolongation being a risk factor for TdP but TdP being the ultimate ventricular arrhythmia *(19, 59, 71)*, since a considerable number of drugs cause QT interval prolongation but are not torsadogenic *(152)*. The relationship between hERG and QT prolongation is not statistically significant in this work, which is potentially related to QT prolongation not being routinely measured in patients, at least for FAERS, whereas the results by Pollard et al. are derived from regulatory Thorough QT/QTc Studies which are designed to detect this effect *(71)*.

More broadly, the lack of one-to-one translation between *in vitro* and *in vivo* effects in this study corroborates previous findings, for example those that observed a lack of correlations between large-scale *in vitro* bioactivities and toxicity observed in animal studies *(185, 186)*, and those

reporting high false positive rates associated with early screening for targets such as hERG and BSEP *(52, 71, 152)*. The reasons behind the challenge to translate *in vitro* to *in vivo* effects include differences between plasma and tissue concentrations *(22, 31, 33)*, varying protein expression across tissues and disease states *(22)*, and interactions between targets and pathways, such as transporter and ion channels off-setting each other's effects *(52, 152)*. Low PPVs and sensitivities observed here and a low recall or previously reported associations may be related to these factors not being taken into account. After taking into account unbound plasma concentrations in this study, the PPVs and LRs for previously reported associations increased (Figure 4.2), thus it is similarly possible that as more of the above-mentioned limitations are addressed, the concordance between *in vitro* and *in vivo* activities would improve, which would become more clear when more data becomes available and additional factors are taken into account.

4.3.4 Presence of established and novel safety targets in current analysis

Next, the significant targets in the current study were compared with those previously reported. In total, this study considered 104 targets out of which 45 were found to have at least one statistically significant association to an AE (Figure 4.7 and Table 4.3). 30 out of these 45 targets are already included on secondary pharmacology panels that were compiled from previous literature *(19, 20, 62)*, so the majority of significant targets in the current study are established safety targets (Figure 4.7). At the same time, out of the 91 safety targets from literature, data was available for only 40 in the current study, highlighting the lack of publicly available experimental data for previously reported safety targets (Figure 4.7). The remaining 15 targets with significant associations in the current study, among which nine are members of the carbonic anhydrase (CA) family, are not included on published panels *(19, 20, 62)* (Table 4.3). Most of the significant associations with CAs identified in this work originate from the SIDER dataset but CA2 and CA4 appear in both datasets (Table 4.3). Apart from the CAs and one other novel target, namely microtubule-associated protein τ, all of the other six novel targets are additional family members of proteins already included on current panels such as serotonin, dopamine, and adrenergic receptors, e.g. dopamine D3 receptor is not currently included on panels but dopamine D1 is. It can be concluded that the current analysis was not able to consider around half of previously reported safety targets due to lack of data, and the novel targets in this work include targets unrelated to previously reported targets as well as family members of previously reported targets.

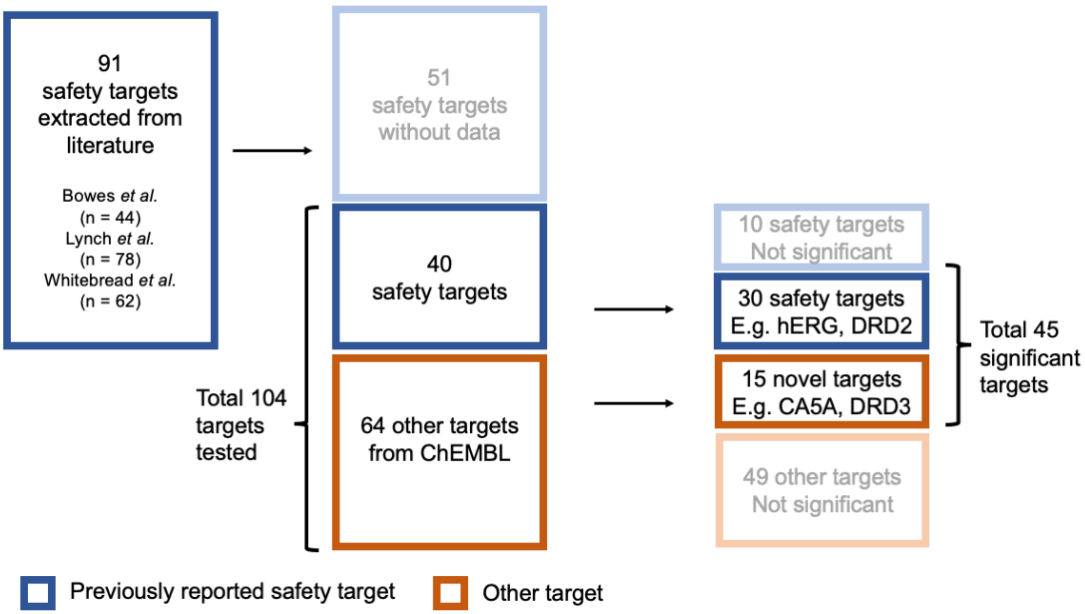

Figure 4.7 Overview of targets in the study, showing the number of targets considered, those that are significantly associated with AEs, and those previously reported on safety target panels. For 51 out of 91 safety targets from literature, no data was available. Of the 45 targets with significant associations to AEs, 30 are listed on current safety target panels whereas 15 are not and thus are potentially novel. CA5B=carbonic anhydrase 5B, DRD2=dopamine D2 receptor, DRD3=dopamine D3 receptor.

With respect to the aim of identifying target-AE associations, the current study is similar to the recent study by Ietswaart et al. *(159)* although their study used a different dataset and methods. None of the 15 potentially novel targets from the current work is among the significant results by Ietswaart et al. *(159)*, but 15 of the human targets previously reported in published panels were retrieved in both the current work and the analysis by Ietswaart et al. *(159)*. In addition, 32 out of 37 human targets with significant associations to AEs in the study by Ietswaart et al. are present on current panels *(19, 20, 62)*. Hence, the results by Ietswaart et al. *(159)* contain a higher fraction of previously reported targets and also identified a different set of potentially novel targets. One potential reason for these differences is that the data used by Ietswaart et al. *(159)* is based on screening data by Eurofins, which may focus on safety targets more than the ChEMBL data underlying the current study does, and may have already excluded redundancy in terms of multiple target family members. In conclusion, the overall set of targets with significant relationships in this work differs from a recent study with similar aims but different methods and underlying data. The next section will consider the specific AEs from this work in more detail.

Table 4.3 All targets in the study that have at least one statistically associated AE. The target class is the second level of the ChEMBL target hierarchy if available, otherwise the first level.

Target	**Target Class**	**Dataset**	**Previously reported** *(19, 20, 62)*
Dopamine D5 receptor	Family A G protein-coupled receptor	FAERS	No
Dopamine D4 receptor	Family A G protein-coupled receptor	FAERS	No
Microtubule-associated protein τ	Unclassified protein	FAERS	No
Dopamine D3 receptor	Family A G protein-coupled receptor	Both	No
α-1d adrenergic receptor	Family A G protein-coupled receptor	Both	No
Carbonic anhydrase 2	Lyase	Both	No
Carbonic anhydrase 4	Lyase	Both	No
Serotonin 6 (5-HT6) receptor	Family A G protein-coupled receptor	SIDER	No
Carbonic anhydrase 14	Lyase	SIDER	No
Carbonic anhydrase 9	Lyase	SIDER	No
Carbonic anhydrase 5B	Lyase	SIDER	No
Carbonic anhydrase 1	Lyase	SIDER	No
Carbonic anhydrase 12	Lyase	SIDER	No
Carbonic anhydrase 9	Lyase	SIDER	No
Carbonic anhydrase 5A	Lyase	SIDER	No
α-1a adrenergic receptor	Family A G protein-coupled receptor	FAERS	Yes
μ opioid receptor	Family A G protein-coupled receptor	FAERS	Yes
α -1b adrenergic receptor	Family A G protein-coupled receptor	FAERS	Yes
Cyclooxygenase-2	Oxidoreductase	FAERS	Yes
Cyclooxygenase-1	Oxidoreductase	FAERS	Yes

Angiotensin-converting enzyme	Protease	FAERS	Yes
Serotonin transporter	Electrochemical transporter	Both	Yes
Serotonin 2a (5-HT2a) receptor	Family A G protein-coupled receptor	Both	Yes
Serotonin 2b (5-HT2b) receptor	Family A G protein-coupled receptor	Both	Yes
Histamine H1 receptor	Family A G protein-coupled receptor	Both	Yes
α -2c adrenergic receptor	Family A G protein-coupled receptor	Both	Yes
Serotonin 2c (5-HT2c) receptor	Family A G protein-coupled receptor	Both	Yes
κ opioid receptor	Family A G protein-coupled receptor	Both	Yes
δ opioid receptor	Family A G protein-coupled receptor	Both	Yes
α-2b adrenergic receptor	Family A G protein-coupled receptor	Both	Yes
Dopamine D2 receptor	Family A G protein-coupled receptor	Both	Yes
Serotonin 7 (5-HT7) receptor	Family A G protein-coupled receptor	Both	Yes
hERG	Voltage-gated ion channel	Both	Yes
Dopamine transporter	Electrochemical transporter	SIDER	Yes
Norepinephrine transporter	Electrochemical transporter	SIDER	Yes
Muscarinic acetylcholine receptor M1	Family A G protein-coupled receptor	SIDER	Yes
Dopamine D1 receptor	Family A G protein-coupled receptor	SIDER	Yes
Muscarinic acetylcholine receptor M4	Family A G protein-coupled receptor	SIDER	Yes
Muscarinic acetylcholine receptor M3	Family A G protein-coupled receptor	SIDER	Yes

Muscarinic acetylcholine receptor M2	Family A G protein-coupled receptor	SIDER	Yes
Type-1 angiotensin II receptor	Family A G protein-coupled receptor	SIDER	Yes
α-2a adrenergic receptor	Family A G protein-coupled receptor	SIDER	Yes
β-1 adrenergic receptor	Family A G protein-coupled receptor	SIDER	Yes
Muscarinic acetylcholine receptor M5	Family A G protein-coupled receptor	SIDER	Yes
σ opioid receptor	Membrane receptor	SIDER	Yes

4.3.5 Target-AE associations with the highest value-added Positive Predictive Values

To examine the individual target-AE associations with the highest PPVs, Table 4.4 shows the significant associations with the highest value-added PPVs in SIDER and FAERS. The table focuses on events within the SOCs of 'nervous system disorders', 'vascular disorders', 'cardiac disorders', and 'respiratory, thoracic and mediastinal disorders', because their link to the function of vital organs makes these a high priority in drug safety *(19, 64)*, and 'hepatobiliary disorders', which is a leading cause of clinical attrition *(187)*. The most highly ranked associations in these categories will be discussed in more detail. Full details of the most highly ranked associations in any organ system are presented in Appendix 1 for FAERS and Appendix 2 for SIDER.

Based on SIDER, the most predictive target-AE association is between CA5B and cholestatic jaundice, a liver disorder, with a value-added PPV of 0.73 (Table 4.4). The CA family of enzymes are involved in acid-base balance and CA inhibitors are used clinically in a range of conditions including ocular disorders, oedema, and seizures, and they are often not entirely selective across CAs *(28, 188–190)*. The association between CA5B and cholestatic jaundice is plausible due to its high tissue expression in the mitochondria *(191)*, which is relevant to liver toxicity *(192)*. While no previous reports of a direct link between CAs and cholestatic jaundice were found in the literature, CA4 and CA12 were identified as protein interactors of targets associated with cholestatic jaundice in the study by Duran-Frigola and Aloy, suggesting the existence of additional indirect links via the protein-protein interaction network *(151)*.

The next most predictive associations concern various muscarinic acetylcholine receptors. In this study, the muscarinic receptor M2 is associated with somnolence (value-added PPV=0.68, Table 4.4), which is in line with acetylcholine being important for wakefulness and the expected effects of M2 antagonism *(193)*. In practice, somnolence is a common side effect of muscarinic acetylcholine M2 and M3 receptor antagonists such as oxybutynin and tolterodine *(194–196)*. Next, the muscarinic acetylcholine receptors M3, M5, M1 and M2 are each associated with tremor with similar value-added PPVs around 0.67 (Table 4.4). The link between the muscarinic acetylcholine receptor M2 and tremor has been previously reported *(19)*, and the links of multiple muscarinic receptors in the current study could either be due to multiple targets being biologically related to the effect, or compound promiscuity. To visualise underlying promiscuity in more detail, Figure 4.8 shows the extent of overlap in active drugs between targets, showing that muscarinic receptors have a high level of overlap in active ligands between each family member, i.e. between half and all active ligands are shared. Thus, it is possible that multiple associations of tremor with muscarinic receptors is due to this promiscuity. In contrast to the CAs, the muscarinic M1 and M2 receptors are included on all published safety panels considered, whilst the muscarinic M5 is included on the Lynch panel *(19, 20, 62)*. In conclusion, the most predictive target-AE associations based on SIDER include a novel link for CA5B, while other associations are previously reported and in line with current mechanistic knowledge and previous literature. However, several links may be driven by compound promiscuity, which cannot be distinguished from independent biological mechanisms based on the current statistical analysis.

Table 4.4. Most highly ranked associations between activity at a protein and reported AEs in the SIDER and FAERS datasets, sorted by their value-added PPVs, and restricted to 'nervous system disorders', 'hepatobiliary disorders', 'cardiac disorders', 'vascular disorders', and 'respiratory, thoracic and mediastinal disorders'.

Target	Adverse event (AE) (MedDRA PT)	System Organ Class	Positive Predictive Value (PPV)	Fraction of drugs with AE	Value-added PPV	Likelihood Ratio	Previously reported in one of *(19, 20, 62)*	Fraction of AE-associated drugs that are active	Number of drugs showing AE
SIDER									
Carbonic anhydrase 5B	JAUNDICE CHOLESTATIC	Hepatobiliary disorders	0.83	0.10	0.73	42.8	No	0.25	20
Muscarinic acetylcholine receptor M3	TREMOR	Nervous system disorders	0.88	0.19	0.68	29.7	No	0.07	96
Muscarinic acetylcholine receptor M2	SOMNOLENCE	Nervous system disorders	1.00	0.32	0.68	inf	No	0.04	180
Muscarinic acetylcholine receptor M5	TREMOR	Nervous system disorders	0.88	0.20	0.67	27.7	No	0.10	72

Muscarinic acetylcholine receptor M1	TREMOR	Nervous system disorders	0.88	0.20	0.67	27.5	No	0.10	72
Muscarinic acetylcholine receptor M2	TREMOR	Nervous system disorders	0.86	0.20	0.66	23.8	Yes *(19, 62)*	0.05	114
FAERS									
Angiotensin-converting enzyme	HYPOVOLAEMIC SHOCK	Vascular disorders	0.60	0.03	0.57	43.8	No	0.30	10
δ opioid receptor	RESPIRATORY DEPRESSION	Respiratory, thoracic and mediastinal disorders	0.45	0.04	0.42	20.4	Yes *(62)*	0.26	19
α-1b adrenergic receptor	ORTHOSTATIC HYPOTENSION	Vascular disorders	0.45	0.05	0.40	15.2	Yes *(20, 62)*	0.24	21
hERG	TORSADE DE POINTES	Cardiac disorders	0.47	0.08	0.38	9.5	No	0.17	40
κ opioid receptor	RESPIRATORY DEPRESSION	Respiratory, thoracic and mediastinal disorders	0.42	0.04	0.38	17.0	No	0.26	19

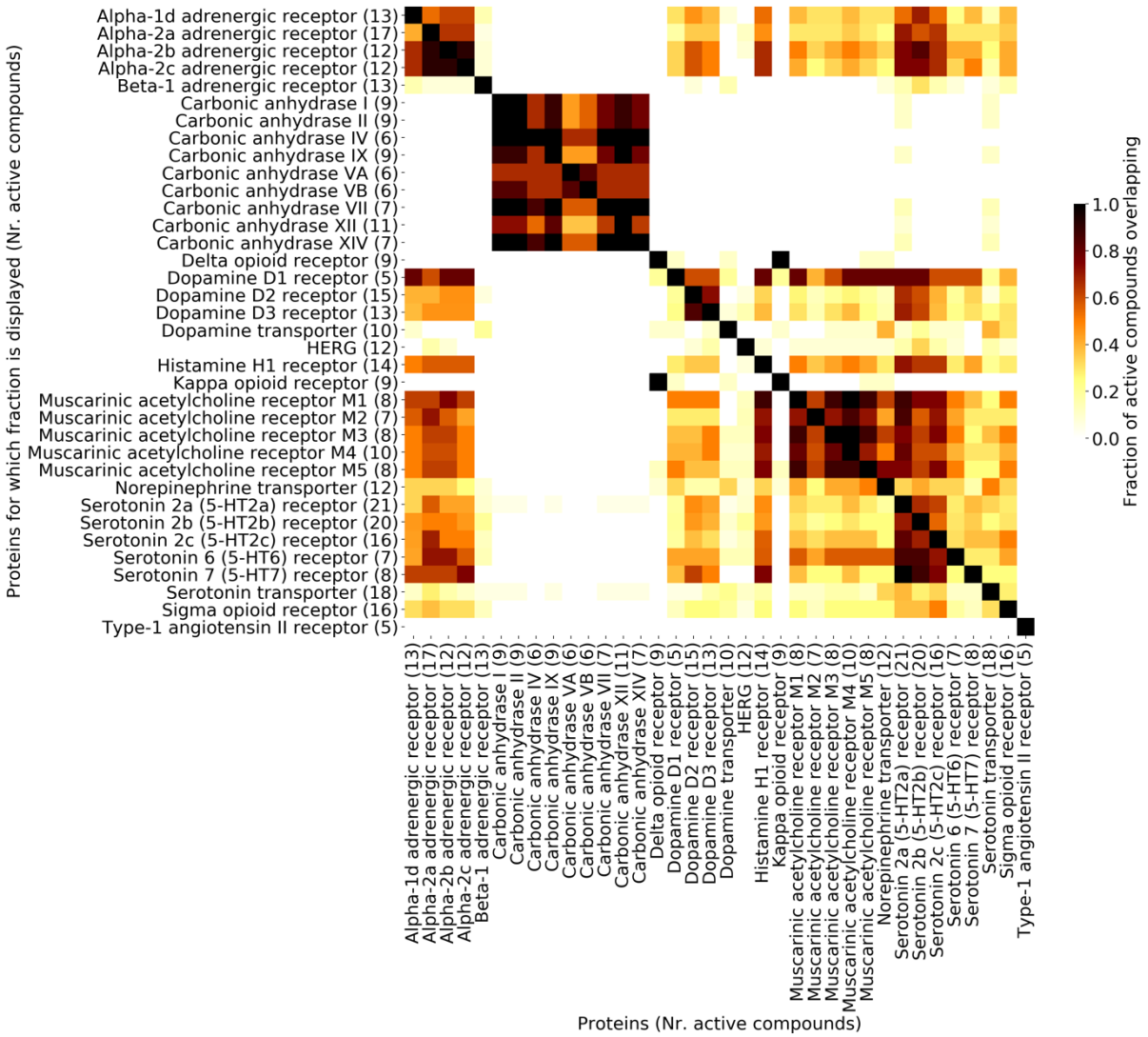

Figure 4.8 Fraction of active drugs, defined by the ratio of the unbound drug plasma concentration over the *in vitro* bioactivity, shared between targets in the SIDER dataset. The fraction corresponds to the target on the y-axis. The number of active compounds at each protein is shown in parentheses. Targets often share active ligands, especially within target families, but also across families. In contrast, the CAs share ligands within the family but hardly so with other targets from other families.

For the analysis of the FAERS database, the association with the highest value-added PPV is between the angiotensin-converting enzyme (ACE), a target only included on the Lynch panel *(62)*, and hypovolemic shock, which is circulatory failure due to fluid loss *(197)* (value-added PPV=0.57, Table 4.4). This is consistent with the fact that ACE inhibitors interfere with the renin-angiotensin system that normally protects against hypovolemia *(198)*. There are case reports in literature that have attributed hypovolemic shock to ACE inhibitors *(199)*.

The known associations between the δ opioid receptor and respiratory depression *(62)* and between the α-1b adrenergic receptor and orthostatic hypotension *(62)* were successfully retrieved in the current analysis with value-added PPVs of 0.42 and 0.40 respectively (Table 4.4). The δ opioid receptor is included on all considered safety panels and the α-1b adrenergic receptor on the Whitebread and Lynch panels *(20, 62)*. Similarly, the association between hERG, one of most studied and screened safety targets *(19, 71, 152)*, and torsade de pointes (TdP) is retrieved with a value-added PPV of 0.38 (Table 4.4). Only QT prolongation is listed among the compiled previously reported associations *(19, 20, 62)* and it has a different MedDRA HLT (investigations) as TdP (ventricular arrhythmias and cardiac arrest), resulting in the association between hERG-TdP in the current work not being annotated as previously reported (Table 4.4). This highlights the challenges in using medical terminologies on a large scale, since descriptions of biological effects may not directly match all related AE terms *(47)*.

The next association in the results is between respiratory depression and the κ opioid receptor (value-added PPV=0.38), a target included on the Bowes and Lynch panels *(19, 62)*. While activation of μ opioid and δ opioid receptors causes respiratory depression *(62, 200)*, the κ opioid receptor is believed to lack this effect *(200)*. However, the presence of this association can also be explained by compound promiscuity, given the high overlap of shared ligands between the opioid receptors in the current study (Figure 4.9). This common profile of active ligands shared between the κ opioid and δ opioid receptors was also identified in the study of secondary pharmacology panels by Bendels et al. *(65)*, which suggested deselection of the δ opioid in an optimised screening panel based on the higher hit rate at the κ opioid receptor. In conclusion, the most predictive associations in FAERS are supported by previous literature or appear to be related to compound promiscuity.

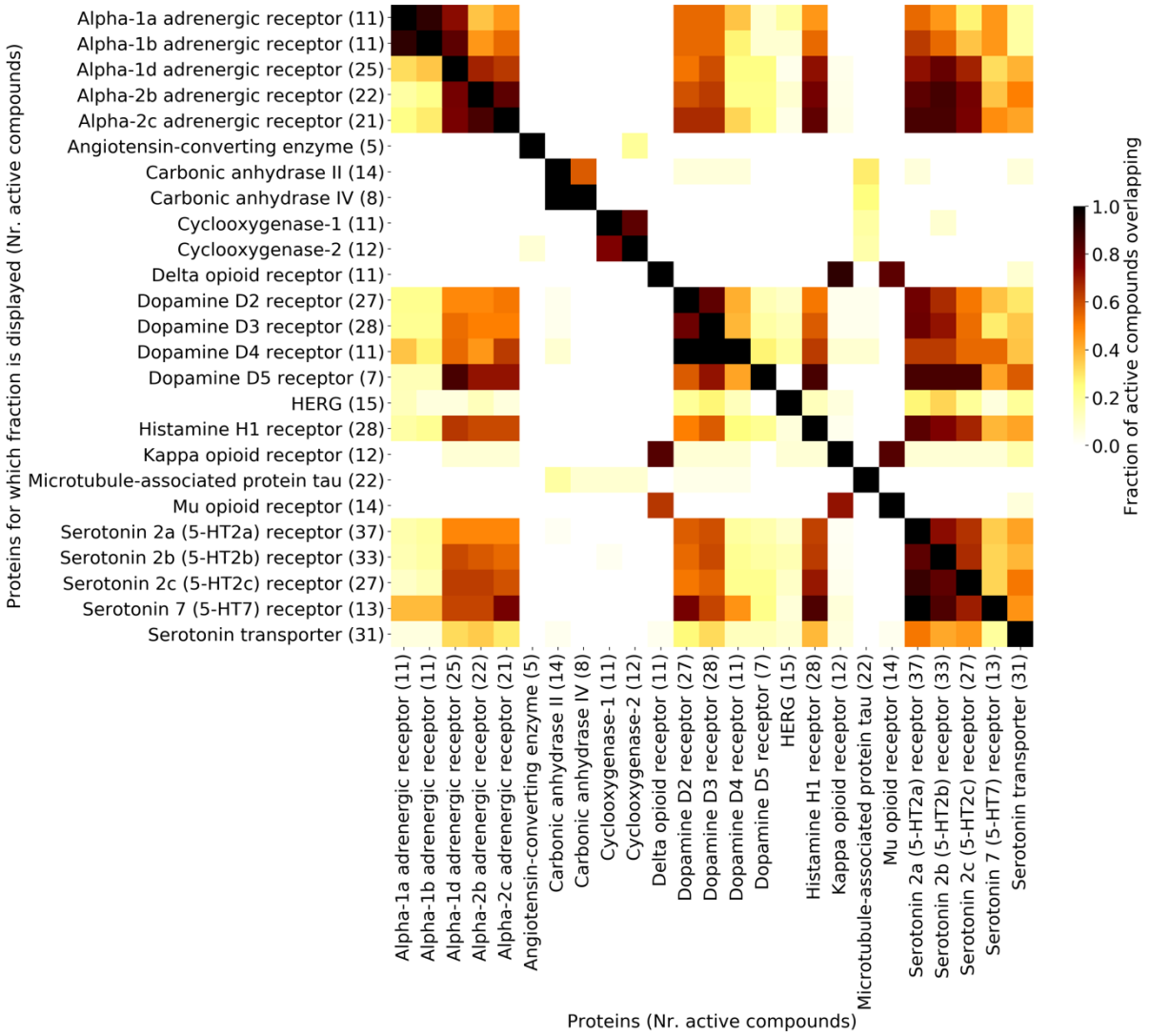

Figure 4.9 Fraction of active drugs, defined by the ratio of the unbound drug plasma concentration over the *in vitro* bioactivity, shared between targets in the FAERS dataset. The fraction corresponds to the target on the y-axis. The number of active compounds at each protein is shown in parentheses. Targets often share active ligands, both within families, e.g. serotonin receptors, and across families, e.g. between serotonin and adrenergic receptors.

4.3.6 Adverse events associated with novel targets and potential value of these associations

To analyse the potential of novel associations from the current study to provide additional information to predict AEs, the novel targets without family members on currently screened secondary pharmacology panels are considered first; these include nine carbonic anhydrases and microtubule-associated protein τ. Table 4.5 lists those targets with significant associations to AEs within the high-priority SOCs mentioned earlier. The most prominent are the CAs which are

associated to a range of AEs in addition to the link between CA5B-cholestatic jaundice, which already appeared in the most highly ranked associations discussed earlier. All the effects associated with CAs are unique to this target family, meaning no other targets are associated to the same effects in the current study. Based on SIDER, CA5A (PPV=0.56), CA12 (PPV=0.45) and CA9 (PPV=0.26) are also associated to cholestatic jaundice (Table 4.5), of which CA5A is most plausible because of its high liver expression *(201)*. Furthermore, CA9 (PPV=0.31) and CA12 (PPV=0.24) are associated with hepatic necrosis (Table 4.5), adding further evidence to the link between CAs and liver effects. CA5A and CA5B are both associated with paraesthesia (PPV=1.0, Table 4.5), which is listed as a side effect of CA inhibitors and has been suggested to be caused by CA activity *(28, 202)*. Based on the FAERS analysis, CA4 is associated with hyperammonaemic encephalopathy (PPV=0.38, Table 4.5), which is consistent with mechanistic knowledge of CA inhibitors on ammonia balance *(28)*. Lastly, CA2 is associated with simple partial seizures, and CA5B with pulmonary oedema (Table 4.5), but both are most likely examples of indication bias in FAERS because these are indications for CA inhibitors *(28, 190)*. This reflects the observation made in chapter 2 that while biases may be reduced by methods such as PSM, they are not fully removed and this is useful to consider when interpreting results such as these. Overall, since none of the existing panels include any members of the CA family and all the associated AEs are unique to this family, the results from this study suggest that CAs might be able to extend the coverage of future safety target panels.

The other target without family members on existing panels is microtubule-associated protein τ, which is associated to liver injury with a PPV=0.28 based on FAERS (Table 4.5). This protein is associated with neurotoxicity and is currently not the therapeutic target of any approved drug. Normal phosphorylation of microtubule-associated protein τ is disturbed by the microcystin group of bacterial toxins, which are also associated with hepatotoxicity *(203)*, potentially providing some support for the observation from this work. Thus, this link could also be a novel association of interest for safety screening.

Next, novel targets that are family members of established safety targets are considered, which include the dopamine D3 receptor, dopamine D4 receptor, α-1d adrenergic receptor, and 5-HT6 receptor. To examine whether these could provide information beyond currently screened targets, the strength of association to AEs was compared with that of the currently screened target and the same AE. For example, the 5-HT6 receptor, not currently included on any of the considered panels, is associated with tremor (PPV= 0.71), but as discussed in the previous section, the

muscarinic M3, M5 and M1 receptors are associated with the same effect with higher PPVs of 0.88 (Table 4.4). Thus, it is possible that these novel associations, such as 5-HT6-tremor, do not provide additional information beyond currently screened targets. The only exception to this pattern is the α-1d adrenergic receptor's association to loss of consciousness with a PPV of 0.54 based on SIDER, an AE to which no other target is associated in the study. In all other cases, novel targets had lower or comparable PPVs to a currently screened target. This shows that different targets can provide similar levels of information about the same AE and might be redundant in the context of safety target screening if there is a large overlap in active drugs across targets. However, the extent of promiscuity will determine whether novel targets can provide additional information to improve the detection of AE-associated drugs, which is explored next.

Table 4.5 Significant associations of novel targets without family members on current panels with AEs in the SOCs 'nervous system disorders', 'hepatobiliary disorders', 'cardiac disorders', and 'respiratory, thoracic and mediastinal disorders'.

Target	Adverse Event (AE)	System Organ Class	Positive Predictive Value (PPV)	Fraction of drugs with AE	Value-added PPV	Likelihood Ratio	Fraction of AE-associated drugs that are active	Number of drugs showing AE	Dataset
Carbonic anhydrase 5B	JAUNDICE CHOLESTATIC	Hepatobiliary disorders	0.83	0.1	0.73	42.8	0.25	20	SIDER
Carbonic anhydrase 5A	PARAESTHESIA	Nervous system disorders	1	0.39	0.61	inf	0.07	85	SIDER
Carbonic anhydrase 5B	PARAESTHESIA	Nervous system disorders	1	0.42	0.58	inf	0.07	80	SIDER
Carbonic anhydrase 5A	JAUNDICE CHOLESTATIC	Hepatobiliary disorders	0.67	0.1	0.56	17.6	0.18	22	SIDER
Carbonic anhydrase 12	JAUNDICE CHOLESTATIC	Hepatobiliary disorders	0.55	0.1	0.45	11.4	0.21	28	SIDER
Carbonic anhydrase 5B	PULMONARY OEDEMA	Respiratory, thoracic and	0.5	0.07	0.43	13.7	0.23	13	SIDER

		mediastinal disorders							
Carbonic anhydrase 4	HYPERAMMONAEMIC ENCEPHALOPATHY	Nervous system disorders	0.38	0.02	0.36	36.6	0.6	5	FAERS
Carbonic anhydrase 9	HEPATIC NECROSIS	Hepatobiliary disorders	0.33	0.02	0.31	19.9	0.38	8	SIDER
Microtubule-associated protein τ	LIVER INJURY	Hepatobiliary disorders	0.36	0.08	0.28	6.2	0.73	11	FAERS
Carbonic anhydrase 9	JAUNDICE CHOLESTATIC	Hepatobiliary disorders	0.33	0.07	0.26	10.5	0.17	23	SIDER
Carbonic anhydrase 12	HEPATIC NECROSIS	Hepatobiliary disorders	0.27	0.03	0.25	13.4	0.38	8	SIDER
Carbonic anhydrase 2	SIMPLE PARTIAL SEIZURES	Nervous system disorders	0.21	0.01	0.21	30.5	0.5	6	FAERS

4.3.7 Considering activity against multiple protein targets for the same AE improves the detection of AE-associated drugs

To find targets that provide non-redundant information, the detection of AE-associated drugs when using single targets versus using combinations was compared. Where multiple targets are associated to the same event, in 38% (FAERS) and 45% (SIDER) of AEs, considering activity at either one of multiple targets improves the detection of AE-associated drugs. Generally, considering two or three targets associated with the AE leads to a median improvement of 20% (FAERS) and 33% (SIDER) in the detection of AE-associated drugs (Figure 4.10). This comes at the cost of worsening PPVs by a median 16% (FAERS) and 21% (SIDER) (Figure 4.10). Thus, the improvement is greater than the reduction in PPVs. The improvements are due to each of the targets identifying different AE-associated drugs.

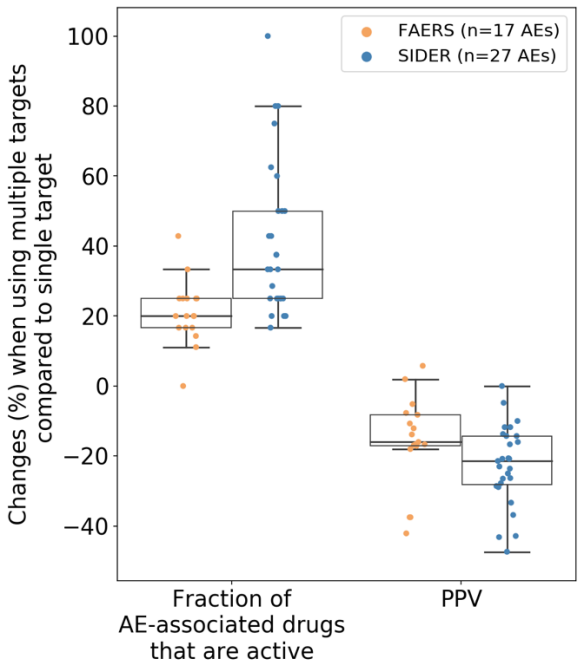

Figure 4.10 Changes in the fraction of AE-associated drugs active at the target ('detected') and PPV when comparing combinations of targets – i.e. activity at either target – for a given AE to the individual targets. Considering activity at multiple targets improves the detection of AE-associated drugs considerably, at a cost of decreasing PPV. Thus, considering activity against multiple targets can help anticipate AEs.

The combination with the greatest improvement in the detection rate compared to the single target is the combination of the β-1 adrenergic receptor and CA5B in relation to orthostatic hypotension in case of the SIDER dataset (Figure 4.11A). Both targets individually detect 12% of AE-associated

drugs, but this is increased to 25% when considering activity at either one of the targets (Figure 4.11A). It is not surprising that a CA is involved in the best performing target set, since active drugs at the CAs generally overlap little with those against other targets in the study, as shown in the promiscuity analysis earlier (Figure 4.8).

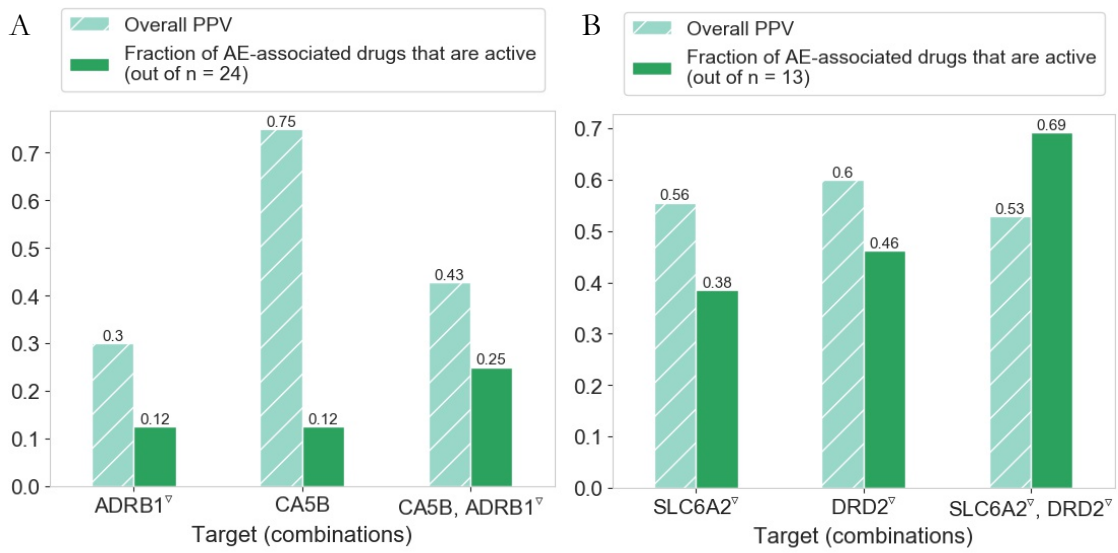

Figure 4.11) PPV and fraction of drugs associated with orthostatic hypotension that are active at the β-1 adrenergic receptor (ADRB1) and carbonic anhydrase 5B (CA5B), as individual targets and in combination, based on 143 drugs that were measured consistently at both targets. (**B**) PPV and fraction of drugs associated with Neuroleptic Malignant Syndrome that are active at the dopamine D2 receptor (DRD2) and norepinephrine transporter (SLC6A2), as individual targets and in combination, based on 320 drugs that were measured at both targets.

The AE with the highest overall percentage of AE-associated drugs detected by a combination of proteins is Neuroleptic Malignant Syndrome in the SIDER dataset (Figure 4.11B); considering activity at either the dopamine D2 receptor or the norepinephrine transporter detects 69% of AE-associated drugs with an overall PPV of 0.32. These targets are consistent with currently known mechanisms behind Neuroleptic Malignant Syndrome *(204)*. The finding that more AE-associated drugs can be identified by using combinations of targets as opposed to single targets is consistent with the knowledge that AEs or toxicities can be caused by multiple mechanisms involving different targets. Thus, this partially explains the low sensitivity of single targets observed earlier (Figure 4.6). Although here only combinations of targets with the 'or' operator were explored, *simultaneous* modulation of targets – 'and' operator – may also explain some AEs, but would not have been discovered in this study.

The analysis performed here to identify targets that provide orthogonal, as opposed to redundant information, bears similarity to a recent approach used by Bendels et al. *(65)*, who used iterative pairwise comparisons of active ligands at different targets to identify redundant versus 'diverse' assays. In the study, assays with many common active ligands were deselected for an optimised panel, whereas targets with low levels of common actives were prioritised for inclusion *(65)*. Their work identified targets sharing a common profile of active ligands, prompting deselection of one of the targets, whereas the combinations in this work highlight the opposite: targets providing complementary information.

In conclusion, in about 40% of AEs in the study, considering activity at each of a set of targets associated with the same event can improve the detection of AE-associated drugs by a median one-fifth to one-third, with in general lower decreases in PPV, indicating a useful trade-off.

4.3.8 Types of AEs associated with protein activity and hence potentially detectable from safety pharmacology screens

In order to estimate what fraction of AEs may be detectable from protein-based safety pharmacology screens, the fraction of unique AEs in the dataset that have at least one significantly associated target was determined. Of all unique AEs, 8.5% in the SIDER dataset and 2.9% in the FAERS dataset have one or more significantly associated target. These low percentages are partially due to the use of unbound plasma concentrations, which restricted the total amount of data available for analysis, since the percentages for the constant cut-off (pChEMBL ≥ 6) dataset are 44.1% (SIDER) and 19.6% (FAERS). This figure for SIDER is similar to the 51% of AEs being related to targets in the study by Kuhn et al. *(150)*. However, based on the trade-off observed in this work between PPV and fraction of AE-associated drugs detected, the above numbers are a result of trading one for the other, the desirability of which depends on the particular situation.

The fact that the percentage is lower for FAERS could be related to the presence of biases and the noisier nature of FAERS reporting *(7)*. FAERS also contains a greater diversity of AEs related to a similar number of drugs compared to SIDER, which was discussed in chapter 2, and target-AE associations are less likely to be detected when only a few drugs are associated, such as to rare events *(150)*. The fractions of unique AEs that can be statistically related to targets could be considered low compared to the figure of 75% of reported AEs being predictable from pharmacology *(10, 15)*, but these latter estimates are based on the frequency of the AE in the population, whereas the numbers from this study look at unique AEs.

To examine differences in how AEs belonging to different SOC are associated with targets, the fraction of unique AEs that is significantly associated with targets was compared with the fraction of unique AEs in the underlying dataset of drug-AE associations, which includes AEs that could not be statistically related to targets (Figure 4.12). The results show that AEs in some SOCs are more frequently associated with targets than AEs from other classes. For example, the largest percentage difference is observed for AEs in the 'metabolism and nutrition disorders' class, which comprise 3.9% (SIDER) and 2.7% (FAERS) of unique AEs in the underlying datasets, but 9.4% (SIDER) and 10.3% (FAERS) of AEs statistically associated with targets (Figure 4.12). The next largest overrepresented classes are GI, nervous system, and psychiatric disorders in FAERS. These results are similar to prominent SOCs in the recent study by Ietswaart et al., who associated HLGTs of AEs with protein targets using FAERS, since GI and psychiatric disorders were among the largest classes in terms of the corresponding HLGTs being significantly related to activities *(159)*. For SIDER, the next largest overrepresented classes are nervous system, 'blood and lymphatic system', and 'respiratory, thoracic and mediastinal' disorders.

The enrichment of 'metabolism and nutrition disorders' and GI disorders suggest these are more often effects that can be related to pharmacological action, as suggested by familiar examples being retrieved in this study, such as cyclooxygenase-1 and gastric ulceration *(19)*, and muscarinic acetylcholine receptor M3-mediated dry mouth and constipation *(19)*. In chapter 2, GI disorders were also found to have the largest number of drugs associated with them (Figure 2.17), thus forming a larger dataset for statistical discovery. AEs in some of the above enriched SOCs also ranked highly for correspondence to target phenotypes in the study by Deaton et al. *(64)*, such as platelet disorders and nonhaemolytic anaemias (blood and lymphatic system), and glucose metabolism disorders (metabolism and nutrition disorders), supporting, from a different angle, the observation that AEs in some SOCs may be more easily related to knowledge of the secondary pharmacological targets than others.

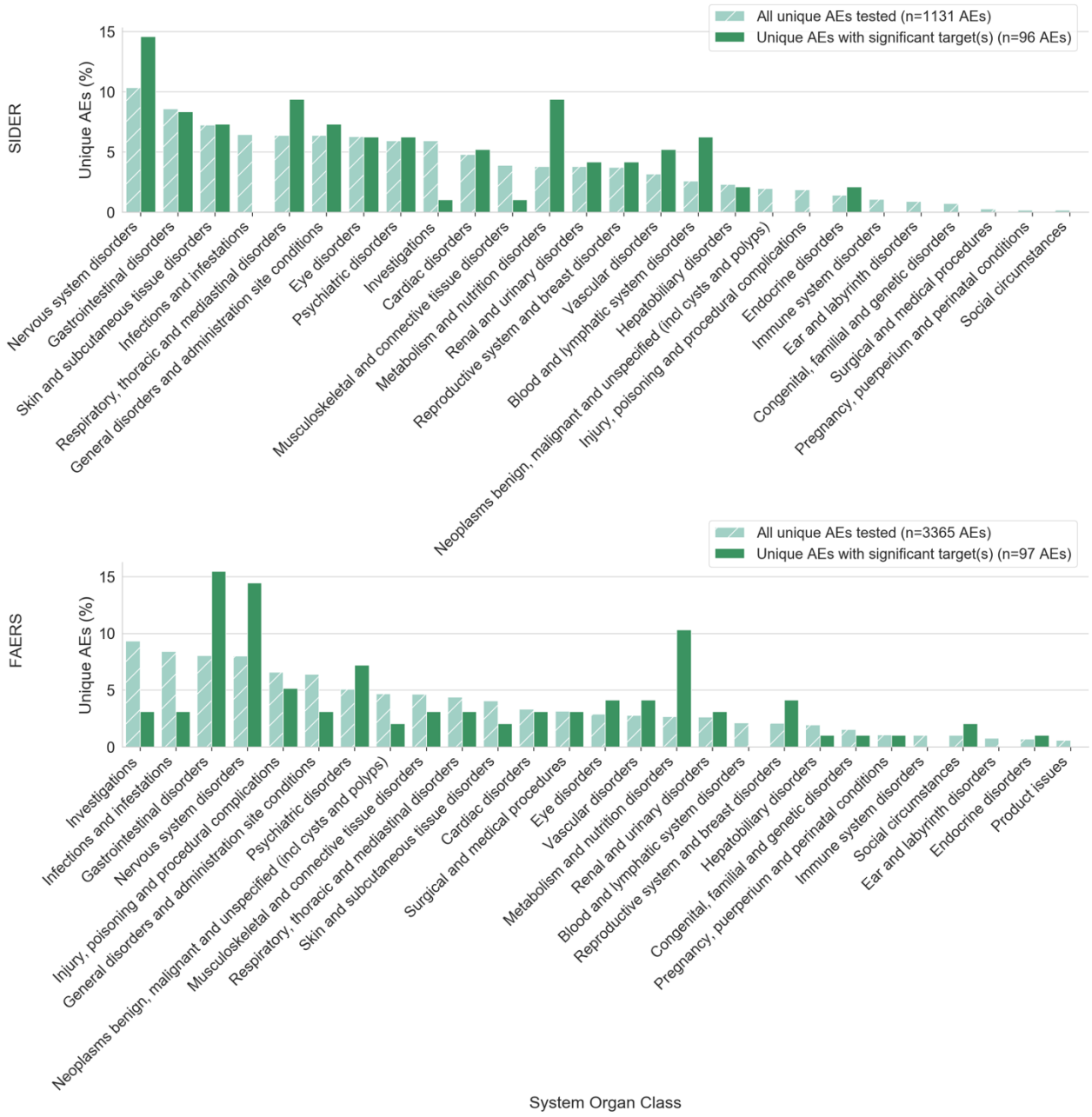

Figure 4.12 Relating AEs to targets by SOC, comparing the AEs in the underlying drug-AE datasets for SIDER and FAERS to those AEs that are statistically associated with targets in the study. AEs in some classes (e.g. 'metabolism and nutrition disorders') are more often associated with targets, whereas AEs in other classes are present in the dataset but rarely associated with targets (e.g. 'infections and infestations').

The overrepresentation of nervous system, psychiatric, and respiratory disorders in both FAERS and SIDER noted above could be related to the prominent presence of GPCRs in the dataset, which frequently target these organ systems and neurotransmission generally *(18, 19)*. In contrast, the largest underrepresented class is 'investigations', which makes up 5.9% (SIDER) and 9.3%

(FAERS) of AEs in the underlying dataset, but only 1.0% (SIDER) and 3.1% (FAERS) of the set related to targets (Figure 4.12). AEs in this category are sometimes relatively unspecific such as 'blood test abnormal', which might be a reason for the lack of associations to targets. Similarly, the next most underrepresented classes are 'musculoskeletal and connective tissue disorders' for SIDER and 'infections and infestations' in FAERS. Many mechanisms by which drugs can increase susceptibility to infections, such as immunosuppression due to cytotoxicity *(14)*, or disruption of the gut microbiota *(205)*, are not covered by *in vitro* pharmacology, explaining their underrepresentation. Overall, it can be concluded that target-AE associations are not uniformly distributed across SOCs, with classes such as 'metabolism and nutrition disorders' and nervous system disorders being most frequently related to pharmacological targets. SOCs with few significant relationships to targets may represent areas in which there is limited *in vitro-in vivo* concordance based on pharmacological activities, thus these may not be expected to be detectable from secondary pharmacology screening.

4.3.9 Protein activities are frequently associated with different AEs in FAERS and SIDER

To investigate the extent to which FAERS and SIDER are complementary for identifying target-AE associations, the overlap in AEs associated with the same targets in either dataset was calculated (Table 4.6). The highest overlap in unique PTs between the datasets is 7% for the dopamine D2 receptor (Table 4.6). Considering the HLT increases the overlap to some extent, but apart from hERG, which is associated with 'ventricular arrhythmias and cardiac arrest' in both datasets, leading to 100% overlap in HLTs, the next highest overlap is still low at only 10% for the dopamine D2 receptor (Table 4.6). This means that in the current work, the FAERS and SIDER datasets contain nearly completely disjointed sets of AEs for the same targets, thus being highly complementary. The complementarity could reflect different AEs being reported in clinical trials versus post-marketing phases, for example due to more detailed patient observation in clinical trials, or due to or differences in short and long-term drug effects. The latter is also supported by the observation that FAERS is the only dataset providing target-AE associations in some SOCs, such as neoplasms and 'pregnancy, puerperium and perinatal conditions' (Figure 4.12), which could be related to long-term use and the more diverse populations exposed in the post-marketing phase. Other reasons for differences in the reported AEs are that AEs already listed on a drug label are less likely to get reported to post-marketing systems *(21)*, and the inclusion of more uncertain reports in FAERS due to lack of causal evidence and submission by patients as opposed to healthcare professionals *(7, 73)*. Thus, it is concluded that FAERS and SIDER provide different

target-AE associations and could each provide added value for detecting target-AE associations. A previous study solely considered AEs reported in both FAERS and SIDER *(64)*, but the results from this work suggest this approach would potentially result in a large part of the AE space not being considered.

Table 4.6. Targets with significantly associated AEs, showing the number of unique associated AEs and the overlap in those between SIDER and FAERS.

Protein	**Number of unique associated AEs (SIDER)**	**Number of unique associated AEs (FAERS)**	**% Unique AEs (MedDRA Preferred Terms) overlapping**	**% Unique MedDRA High Level Terms overlapping**
Dopamine D2 receptor	31	15	7.0	10.3
Dopamine D3 receptor	13	8	5.0	5.3
Serotonin transporter	6	22	3.7	4.3
Serotonin 2a (5-HT2a) receptor	9	27	2.9	3.2
Carbonic anhydrase 5B	27	0	0.0	0.0
Muscarinic acetylcholine receptor M3	24	0	0.0	0.0
Norepinephrine transporter	19	0	0.0	0.0
α-1d adrenergic receptor	18	4	0.0	0.0
Muscarinic acetylcholine receptor M5	15	0	0.0	0.0
Muscarinic acetylcholine receptor M1	15	0	0.0	0.0
Dopamine D1 receptor	15	0	0.0	0.0
Carbonic anhydrase 5A	15	0	0.0	0.0

Histamine H1 receptor	13	16	0.0	0.0
Serotonin 6 (5-HT6) receptor	11	0	0.0	0.0
Muscarinic acetylcholine receptor M4	10	0	0.0	0.0
β-1 adrenergic receptor	9	0	0.0	0.0
Carbonic anhydrase 12	8	0	0.0	0.0
Carbonic anhydrase 9	8	0	0.0	0.0
σ opioid receptor	8	0	0.0	0.0
Carbonic anhydrase 2	5	7	0.0	0.0
Carbonic anhydrase 1	5	0	0.0	0.0
α-2c adrenergic receptor	4	13	0.0	0.0
Serotonin 2c (5-HT2c) receptor	4	12	0.0	0.0
α-2b adrenergic receptor	3	1	0.0	0.0
Carbonic anhydrase 4	3	1	0.0	0.0
Muscarinic acetylcholine receptor M2	3	0	0.0	0.0
κ opioid receptor	2	1	0.0	0.0
δ opioid receptor	2	1	0.0	0.0
α-2a adrenergic receptor	2	0	0.0	0.0
Carbonic anhydrase 7	2	0	0.0	0.0
Serotonin 7 (5-HT7) receptor	1	20	0.0	0.0
Serotonin 2b (5-HT2b) receptor	1	10	0.0	0.0
hERG	1	1	0.0	100.0
Dopamine transporter	1	0	0.0	0.0

Type-1 angiotensin II receptor	1	0	0.0	0.0
Carbonic anhydrase 14	1	0	0.0	0.0
α-1a adrenergic receptor	0	19	0.0	0.0
Dopamine D4 receptor	0	14	0.0	0.0
μ opioid receptor	0	8	0.0	0.0
α-1b adrenergic receptor	0	7	0.0	0.0
Cyclooxygenase-2	0	6	0.0	0.0
Cyclooxygenase-1	0	6	0.0	0.0
Angiotensin-converting enzyme	0	3	0.0	0.0
Dopamine D5 receptor	0	1	0.0	0.0
Microtubule-associated protein τ	0	1	0.0	0.0

4.4 Conclusions

In this chapter, associations between drugs' pharmacological activities and AEs observed in clinical trials and during post-marketing surveillance were identified and quantified.

Taking into account unbound drug plasma concentrations reduced the size of the dataset by about half, but increased the PPV and LR of associations reported in previous literature at the cost of lower recall. This suggests that adjusting activity calls with unbound plasma concentrations results in more precise associations but that recall is currently limited by data availability.

Many of the most predictive associations in the current study are supported by previous literature, but some associations appeared to be related to compound promiscuity, which remains a source of potential false positive associations in studies of statistical nature. Thus, regarding any statistical associations derived in this study, further research would be needed to confirm mechanistic links. Key novel findings such as the association of CA5A and CA5B with liver effects would be suggested for further investigation and future inclusion in safety target panels.

The global assessment of the distribution of predictive values and sensitivity of target-AE associations shows that *in vitro* bioactivities rarely directly translate into *in vivo* effects based on the datasets studied, with predictive values for individual *in vitro* activities centred around 40%, with lower values (25%) applying to FAERS datapoints. A range of limitations of the study, such as survivor bias in approved drugs and incomplete consideration of dose, as well as of current knowledge and available data, such as tissue distribution of drugs, may be related to these observations. The low sensitivity could be partially rescued by considering activity at multiple targets associated with the same event, which improved the detection of AE-associated drugs in the case of around 40% of AEs at the cost of lower precision.

Furthermore, it was found that significant AE associations are not uniformly distributed across SOCs, and AEs in enriched SOCs such as nervous system and GI disorders are more frequently associated with pharmacological activities in this study, suggesting they may be more easily detectable from secondary pharmacology screening.

Lastly, FAERS and SIDER show a low overlap in terms of AEs and resulting target-AE associations, indicating that each provides unique, complementary information for the study of target-AE associations.

5 Conclusion

5.1 Summary of contribution

The aim of this research was to investigate target-AE associations in a systematic way, using datasets with improved reliability and accuracy compared to previous work. Therefore, confounding factors in the post-marketing drug-AE dataset were controlled with PSM, and drug plasma concentrations were integrated with *in vitro* bioactivity data to improve the *in vivo* relevance of drug-target annotations.

The value in the first part of the work, applying PSM on the FAERS AEOLUS database and examining the impact on indication bias, lies in applying this method for bias control, which originated in the field of epidemiology, to the area of drug discovery informatics, given only few related analyses previously conducted focused on deriving target-AE relationships whilst controlling confounding factors. Extending the work by Tatonetti et al. *(206)*, the current work constitutes an independent validation of the benefit of using PSM in reducing false associations of drugs with drug indications in FAERS.

Next, this work assessed the impact of using unbound drug plasma concentrations on activity calls as well as on derived target-AE associations. While existing literature on safety pharmacology screening frequently stresses the importance of considering unbound drug plasma concentrations, and this being standard practice in risk assessment, this thesis has for the first time brought insights into the impact of doing so on a larger scale using publicly available data. The halving of available drug-target datapoints by requiring drug plasma data, and the lower overall retrieval of previously reported target-AE associations when using adjusted activity calls shows that data availability is still a limitation. Nevertheless, the results suggest that using unbound drug plasma concentrations results in activity calls that are more useful (higher LRs) for anticipating AEs and have higher precision, based on the retrieval of currently known safety targets. Therefore, this work contributes insight into the trade-off between the availability and the precision of currently available data.

The target-AE associations computed in this work provide novel suggestions of targets for future investigation and possible inclusion on future screening panels, such as CA5A and CA5B in relation to liver toxicity. In addition, this work provides a large-scale quantification of target-AE

associations which was previously unavailable, presenting the distribution of LRs and predictive values of diverse target-AE associations across the dataset. These results help to identify the targets most strongly associated with AEs, provide evidence for the design of future safety pharmacology panels and provide suggestions for the development of *in silico* models. The quantified results are useful to the interpretation of secondary pharmacology results by providing, based on retrospective analysis, an estimate of the probability of *in vivo* events given a certain bioactivity.

The global assessment of predictive values of target-AE associations in this work has highlighted that many targets do not have a perfect association to AEs *in vivo*, with AEs in some SOCs being more frequently related to pharmacological effects than those in other SOCs. The value of these results lies in the highlighting of areas in need of future development, such as the lack of complete functional bioactivity data (agonism versus antagonism), and the lack of data regarding the distribution of drugs across different tissues, which may be a factor in the low sensitivities and value-added PPVs observed here. The results are also relevant to developments in predictive toxicology, some of which focus on applying machine learning to predict target bioactivities for lists of toxicity-related targets or assays *(207–209)*. Such efforts will be most useful if the targets modelled are strong indicators of *in vivo* effects.

Lastly, this work prevents novel findings in two other areas, which are (1) the extent to which combinations of targets may be able to increase the detection of AE-associated drugs without adding redundancy to targets included in screening and (2) the extent of complementarity between FAERS and SIDER. Overall, based on a comparison of the the drug-AE datasets and the computed target-AE associations, it was found that FAERS and SIDER have a low overlap of AE terms and provide complementary information.

5.2 Limitations

Several limitations primarily relating to the underlying data have affected this work. Chapter 2 discussed the limitations of the FAERS and SIDER datasets in terms of the compiled drug-AE relationships, such as the biases affecting FAERS. The datasets used in this work are inevitably not fully comprehensive and contain possible false labels. The bioactivity data is affected by sparsity and the inclusion of predicted datapoints with their associated errors, thus it is possible that the absence or inaccuracy of certain associations in this dataset is related to these limitations. Furthermore, there is a lack of comprehensive functional (agonism or antagonism) effects on targets and thus this was not included in the work. This may have resulted in the masking of

associations only associated with certain functional effects or modes of action. Lastly, chapter 3 discussed some of the limitations of the drug plasma concentrations, including that they only consider the therapeutic range and are not specific to the patients affected by AEs in FAERS or SIDER. Overall, these factors affecting the underlying data sources have most likely limited the sensitivity and specificity in terms of discovering underlying, causative target-AE relationships in this study, and maybe have contributed to low predictive values observed.

Regarding data limitations, a theme that emerged throughout the work is the balance between data availability and data quality. For example, attempting to increase the quality of the data by using unbound plasma concentrations resulted in a decrease in the number of drug-target datapoints available. Similarly, when setting minimum thresholds for the target prediction model performance and applicability domain, and a minimum number of drugs in two-by-two tables, less data became available and certain target classes were underrepresented such as kinases. The trade-off between data quality and quantity represents a recurring limitation in this work of which the impact is difficult to evaluate without a gold standard of target-AE associations.

Chapter 4 discussed how the 'survivor bias' affecting marketed drugs, as well as the incomplete consideration of dose, may explain why some previously reported associations were not observed in the current work and other associations may have low predictive values. However, even when additional doses would have been considered, there is also considerable inter-individual variation in pharmacokinetics that can contribute to ADRs *(14, 36)*. This is one of a wider range of factors that were beyond the scope of this study, with others being drug metabolites and patient genetics affecting the susceptibility to ADRs *(14, 36)*.

Ultimately, there will also be a limit to the extent to which currently known or predicted specific drug-target interactions may be able to explain ADRs, because additional mechanisms that cause ADRs, such as immunological reactions *(14, 210)*, are not covered by pharmacological drug-target interactions. Furthermore, there are unexplored proteins in the human proteome which may be relevant to ADRs but are not yet included in currently available experimental data *(211)*. Lastly, the statistical nature of the current work means that the causal nature of suggestions of associations between targets and AEs cannot be confirmed, and further *in vitro* or *in vivo* studies would be needed.

5.3 Future outlook

While the amount of bioactivity data in ChEMBL and other resources continues to increase, given this is dependent on manual extraction from scientific literature *(105)*, it is not likely that more complete matrices of bioactivity data for existing drugs will be available soon. The most obvious ways to increase the amount of bioactivity data for analysis would be to consider resources that have integrated multiple other bioactivity databases such as ExCAPE-DB *(108)*, which by incorporating PubChem includes more data, or by drawing on legacy datasets from Bioprint *(69)* or the pharmaceutical industry *(212)*, although the latter two are not publicly accessible. An alternative approach being developed by Lhasa Limited is to train target prediction models on private data from multiple companies and then make target predictions on public compounds available to other users without sharing the original training data *(213, 214)*. This should result in better target predictions models and more predictions being available, given the larger and potentially more diverse training set, as was demonstrated in the case of hERG *(213, 214)*. This approach could be useful for the study of marketed drugs and would be expected to provide more reliable and a greater number of target predictions, although these would be limited to targets that have been consistently measured at different companies.

A feasible, natural extension of the current work using public domain data would be to consider the Tox21 screening data *(111, 153, 215)*, which has the advantages of being a fully screened matrix that includes marketed drugs, and includes assays for the detection of agonism and antagonism modes as well as a range of more general stress response pathways. This would address some of the limitations of considering pharmacology data only, and potentially expand the range of AEs that may be related to *in vitro* assay activities. Different normalisation procedures have been published by a number of institutes involved in generating the Tox21 dataset *(215–217)*, each emphasizing different aspects, thus consideration needs to be given to the pre-processing of the data before being able to extract the AC_{50} values from the dataset. A few studies have used the dataset to relate targets to different *in vivo* effects *(154, 186)* although more have tried to predict *in vivo* toxicities directly *(111, 218, 219)*.

Increasing the amount of drug plasma concentration data by manual extraction from scientific literature is laborious, but progress is being made in automated approaches for the prediction of drug plasma concentrations, such as the High-Throughput Toxicokinetics (HTTK) package *(220)*. The parameters needed to obtain predictions for the steady-state drug plasma concentrations with

the HTTK package are the *in vitro* intrinsic clearance and the unbound drug fraction, which may be determined using *in vitro* assays *(220)*. The HTTK package includes such data derived from human-specific *in vitro* assays for 94 drugs *(220)*, so additional data gathering would be needed to use the HTTK models for more drugs. While the HTTK models are less sophisticated than physiologically-based pharmacokinetic models based on *in vivo* data *(220)*, and predictions will be less accurate than measured values, it would valuable to examine whether using these high-throughput style predictions are an improvement over using 1 or 10 μM as a cut-off for activity calls in secondary pharmacology screening, which is current practice at early stages *(19, 22, 65)*. A recent study found that using the HTTK package improved concordance between *in vitro* activities in ToxCast and *in vivo* toxicities in rat for a set of 84 chemicals *(221)*. Measured plasma concentrations could be supplemented by predicted concentrations, similar to the way target predictions were used in chapter 3, and the approach could be evaluated by comparing the resulting drug-AE associations, the same way the dataset derived from using a constant cut-off was compared to using adjusted activity calls based on the unbound plasma concentrations in chapter 4.

Another idea for future work is to consider greater variation in plasma concentrations, as opposed to the single, therapeutic concentrations used in this work. Additionally the 'toxic' concentration could be used, which were reported for up to 1,000 drugs and chemicals in the publication by Schulz et al. *(176)* that also provided many therapeutic plasma concentrations used in this work, but it might be time-consuming to obtain this data for a larger set of drugs. However, considering higher plasma concentrations would represent situations such as drug-drug interactions, inter-individual variation, and overdose, which are undoubtedly a factor in some ADRs *(14, 36)*, and would therefore be valuable to consider. At higher plasma concentrations, drugs can interact with more targets, and more AEs may then be related to target activities, which might result in a higher retrieval of previously reported target-AE associations as well as help to identify novel relationships. It would be useful if the extent of pharmacokinetic variation could be represented as a distribution in order to derive different activity calls for different concentrations. The most easily accessible relevant data for this would be the range of effective therapeutic plasma concentrations, which have been reported for sets of marketed drugs *(36, 176)*.

Given the large jump made between *in vitro* and *in vivo* effects in this work, and the many processes existing in between that influence their concordance, an interesting recent development are the concepts of quantitative adverse outcome pathways (qAOPs), which aim to quantitatively link

molecular, cellular and organ biology leading to AEs *(222)* and quantitative systems toxicology (QST), which is defined as "an approach to quantitatively understand the toxic effects of a chemical on a living organism, from molecular alterations to phenotypical observations, through the integration of computational and experimental methods" *(223)*. Only tens of qAOPs have currently been developed *(222)*, and both qAOPs and QST face challenges in data availability *(222, 223)*. Considerable scientific efforts are needed to comprehensively understand and represent toxicity and pharmacological pathways *(33, 222, 223)*, thus this is still years away. However, the better we are ultimately able to represent the multitude of processes involved in drugs' effects and pharmacokinetic behaviour, the better we will be able to translate *in vitro* into *in vivo* effects and make progress in the development of safe medicines.

6 References

1. J. K. Aronson, R. E. Ferner, Clarification of terminology in drug safety, *Drug Saf.* **28**, 851–870 (2005).

2. R. E. Ferner, P. McGettigan, Adverse Drug Reactions, *BMJ* **363**, k4051 (2018).

3. World Health Organization, WHO Collaborating Centre for International Drug Monitoring, The importance of pharmacovigilance: safety monitoring of medicinal products, (2002) (available at https://apps.who.int/iris/bitstream/handle/10665/42493/a75646.pdf;sequence=1, accessed June 29, 2020).

4. European Medicines Agency, Guideline on good pharmacovigilance practices (GVP) - Annex I - Definitions (Rev 4), EMA/876333/2011, (2017) (available at https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-annex-i-definitions-rev-4_en.pdf, accessed June 29, 2020).

5. Electronic Medicines Compendium, Diclofenac Potassium 25 mg Tablets - Summary of Product Characteristics (SmPC), (available at https://www.medicines.org.uk/emc/product/5791/smpc, accessed May 20, 2020).

6. U.S. Food and Drug Administration, Code of Federal Regulations Title 21--Food and Drugs. 21CFR312.32, (2019) (available at https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?fr=312.32, accessed June 29, 2020).

7. T. J. Moore, M. R. Cohen, C. D. Furberg, Serious adverse drug events reported to the Food and Drug Administration, 1998-2005, *Arch. Intern. Med.* **167**, 1752–1759 (2007).

8. J. C. Bouvy, M. L. De Bruin, M. A. Koopmanschap, Epidemiology of adverse drug reactions in Europe: a review of recent observational studies, *Drug Saf.* **38**, 437–453 (2015).

9. L. Hazell, S. A. W. Shakir, Under-reporting of adverse drug reactions, *Drug Saf.* **29**, 385–396 (2006).

10. J. Lazarou, B. H. Pomeranz, P. N. Corey, Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies, *JAMA* **279**, 1200–1205 (1998).

11. M. Pirmohamed, S. James, S. Meakin, C. Green, A. K. Scott, T. J. Walley, K. Farrar, B. K. Park, A. M. Breckenridge, Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients, *BMJ* **329**, 15–19 (2004).

12. M. J. Waring, J. Arrowsmith, A. R. Leach, P. D. Leeson, S. Mandrell, R. M. Owen, G. Pairaudeau, W. D. Pennie, S. D. Pickett, J. Wang, O. Wallace, A. Weir, An analysis of the attrition of drug candidates from four major pharmaceutical companies, *Nat. Rev. Drug Discov.* **14**, 475–486 (2015).

13. D. Cook, D. Brown, R. Alexander, R. March, P. Morgan, G. Satterthwaite, M. N. Pangalos, Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework, *Nat. Rev. Drug Discov.* **13**, 419–431 (2014).

14. A. Lee, Ed., *Adverse Drug Reactions* (Pharmaceutical Press, London, 2nd ed, 2006).

15. W. S. Redfern, I. D. Wakefield, H. Prior, C. E. Pollard, T. G. Hammond, J.-P. Valentin, Safety pharmacology – a progressive approach, *Fundam. Clin. Pharmacol.* **16**, 161–173 (2002).

16. R. H. B. Meyboom, A. C. G. Egberts, I. R. Edwards, Y. A. Hekster, F. H. P. de Koning, F. W. J. Gribnau, Principles of signal detection in pharmacovigilance, *Drug Saf.* **16**, 355–365 (1997).

17. A. Sundström, P. Hallberg, Data mining in pharmacovigilance - Detecting the unexpected, *Drug Saf.* **32**, 419–427 (2009).

18. R. Santos, O. Ursu, A. Gaulton, A. P. Bento, R. S. Donadi, C. G. Bologa, A. Karlsson, B. Al-Lazikani, A. Hersey, T. I. Oprea, J. P. Overington, A comprehensive map of molecular drug targets, *Nat. Rev. Drug Discov.* **16**, 19–34 (2017).

19. J. Bowes, A. J. Brown, J. Hamon, W. Jarolimek, A. Sridhar, G. Waldron, S. Whitebread, Reducing safety-related drug attrition: the use of in vitro pharmacological profiling, *Nat. Rev. Drug Discov.* **11**, 909–922 (2012).

20. S. Whitebread, J. Hamon, D. Bojanic, L. Urban, In vitro safety pharmacology profiling: an essential tool for successful drug development, *Drug Discov. Today* **10**, 1421–1433 (2005).

21. R. Ferner, P. McGettigan, The patient who reports a drug allergy, *BMJ* **368**, l6791 (2020).

22. S. Jenkinson, F. Schmidt, L. Rosenbrier Ribeiro, A. Delaunois, J.-P. Valentin, A practical guide to secondary pharmacology in drug discovery, *J. Pharmacol. Toxicol. Methods* **In Press** (2020), doi:10.1016/j.vascn.2020.106869.

23. D. C. Liebler, F. P. Guengerich, Elucidating mechanisms of drug-induced toxicity, *Nat. Rev. Drug Discov.* **4**, 410–420 (2005).

24. J.-P. Valentin, R. Bialecki, L. Ewart, T. Hammond, D. Leishmann, S. Lindgren, V. Martinez, C. Pollard, W. Redfern, R. Wallis, A framework to assess the translation of safety pharmacology data to humans, *J. Pharmacol. Toxicol. Methods* **60**, 152–158 (2009).

25. MedDRA - The Medical Dictionary for Regulatory Activities, (available at www.meddra.org, accessed April 13, 2020).

26. G. H. Merrill, The MedDRA paradox, *AMIA. Annu. Symp. Proc.* **2008**, 470–474 (2008).

27. MedDRA, MedDRA Hierarchy, (available at https://www.meddra.org/how-to-use/basics/hierarchy, accessed June 29, 2020).

28. J. G. Hardman, L. E. Limbird, A. G. Gilman, *Goodman & Gilman's The Pharmacological Basis of Therapeutics* (McGraw-Hill Medical, New York, 10th ed, 2001).

29. P. Rolan, V. Molnár, in *The Textbook of Pharmaceutical Medicine*, (John Wiley & Sons, Ltd, 2013), pp. 113–131.

30. D. J. Birkett, *Pharmacokinetics Made Easy* (McGraw-Hill, 2nd edition, 2002).

31. P. Y. Muller, M. N. Milton, The determination and interpretation of the therapeutic index in drug development, *Nat. Rev. Drug Discov.* **11**, 751–761 (2012).

32. D. A. Smith, L. Di, E. H. Kerns, The effect of plasma protein binding on in vivo efficacy: misconceptions in drug discovery, *Nat. Rev. Drug Discov.* **9**, 929–939 (2010).

33. S. Polak, Z. Tylutki, M. Holbrook, B. Wiśniowska, Better prediction of the local concentration–effect relationship: the role of physiologically based pharmacokinetics and quantitative systems pharmacology and toxicology in the evolution of model-informed drug discovery and development, *Drug Discov. Today* **24**, 1344–1354 (2019).

34. R. M. Julien, C. D. Advokat, J. E. Comaty, *A Primer of Drug Action* (Worth Publishers, 12th edition, 2011).

35. J. M. Ritter, R. J. Flower, G. Henderson, Y. Kong Loke, D. MacEwan, H. P. Rang, *Rang and Dale's Pharmacology* (Elsevier, Edinburgh, 9th edition, 2020).

36. L. L. Brunton, R. Hilal-Dandan, B. C. Knollmann, Eds., *Goodman & Gilman's The Pharmacological Basis of Therapeutics* (McGraw Hill, New York, 13th ed, 2018).

37. G. A. Holdgate, T. D. Meek, R. L. Grimley, Mechanistic enzymology in drug discovery: a fresh perspective, *Nat. Rev. Drug Discov.* **17**, 115–132 (2018).

38. J. W. Johnson, S. E. Kotermanski, Mechanism of action of memantine, *Curr. Opin. Pharmacol.* **6**, 61–67 (2006).

39. H.-S. V. Chen, S. A. Lipton, The chemical biology of clinically tolerated NMDA receptor antagonists, *J. Neurochem.* **97**, 1611–1626 (2006).

40. M. P. Gleeson, A. Hersey, D. Montanari, J. Overington, Probing the links between in vitro potency, ADMET and physicochemical parameters, *Nat. Rev. Drug Discov.* **10**, 197–208 (2011).

41. A. C. Pan, D. W. Borhani, R. O. Dror, D. E. Shaw, Molecular determinants of drug–receptor binding kinetics, *Drug Discov. Today* **18**, 667–673 (2013).

42. F. J. Ehlert, *Affinity and Efficacy: The Components of Drug-receptor Interactions* (World Scientific, New Jersey, 2014).

43. R. Z. Cer, U. Mudunuri, R. Stephens, F. J. Lebeda, IC50-to-Ki: a web-based tool for converting IC50 to Ki values for inhibitors of enzyme activity and ligand binding, *Nucleic Acids Res.* **37**, W441–W445 (2009).

44. T. Kalliokoski, C. Kramer, A. Vulpetti, P. Gedeck, Comparability of mixed IC50 data – a statistical analysis, *PLoS ONE* **8** (2013).

45. J.-U. Peters, J. Hert, C. Bissantz, A. Hillebrecht, G. Gerebtzoff, S. Bendels, F. Tillier, J. Migeon, H. Fischer, W. Guba, M. Kansy, Can we discover pharmacological promiscuity early in the drug discovery process?, *Drug Discov. Today* **17**, 325–335 (2012).

46. X. Jalencas, J. Mestres, On the origins of drug polypharmacology, *MedChemComm* **4**, 80–87 (2012).

47. L. Urban, V. F. Patel, R. J. Vaz, Eds., *Antitargets and Drug Safety* (Wiley-VCH, Weinheim, Germany, 2015).

48. J.-U. Peters, Polypharmacology – Foe or friend?, *J. Med. Chem.* **56**, 8955–8971 (2013).

49. A. A. Antolin, P. Workman, J. Mestres, B. Al-Lazikani, Polypharmacology in precision oncology: current applications and future prospects, *Curr. Pharm. Des.* **22**, 6935–6945 (2016).

50. B. L. Roth, D. J. Sheffler, W. K. Kroeze, Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia, *Nat. Rev. Drug Discov.* **3**, 353–359 (2004).

51. A. Anighoro, J. Bajorath, G. Rastelli, Polypharmacology: challenges and opportunities in drug discovery, *J. Med. Chem.* **57**, 7874–7887 (2014).

52. E. A. G. Blomme, Y. Will, Toxicology strategies for drug discovery: present and future, *Chem. Res. Toxicol.* **29**, 473–504 (2016).

53. Y. J. Ribeill, in *The Textbook of Pharmaceutical Medicine*, (2013), pp. 1–31.

54. J. Posner, S. Warrington, in *The Textbook of Pharmaceutical Medicine*, (John Wiley & Sons, Ltd, 2013), pp. 143–154.

55. L. Müller, E. Husar, in *The Textbook of Pharmaceutical Medicine*, (John Wiley & Sons, Ltd, 2013), pp. 42–81.

56. S. F. Hobbiger, B. Patel, E. Swain, in *The Textbook of Pharmaceutical Medicine*, (John Wiley & Sons, Ltd, 2013), pp. 235–253.

57. R. J. Vaz, T. Klabunde, Eds., *Antitargets* (Wiley-VCH, Weinheim, Germany, 2008).

58. Rothman Richard B., Baumann Michael H., Savage Jason E., Rauser Laura, McBride Ace, Hufeisen Sandra J., Roth Bryan L., Evidence for possible involvement of 5-HT2B receptors in the cardiac valvulopathy associated with fenfluramine and other serotonergic medications, *Circulation* **102**, 2836–2841 (2000).

59. M. L. De Bruin, M. Pettersson, R. H. B. Meyboom, A. W. Hoes, H. G. M. Leufkens, Anti-HERG activity and the risk of drug-induced arrhythmias and sudden death, *Eur. Heart J.* **26**, 590–597 (2005).

60. D. G. Brown, G. F. Smith, H. J. Wobst, Promiscuity of in vitro secondary pharmacology assays and implications for lead optimization strategies, *J. Med. Chem.* **63**, 6251–6275 (2020).

61. SafetyScreen44 Panel, Cerep, (available at https://www.eurofinsdiscoveryservices.com/catalogmanagement/viewitem/SafetyScreen44-Panel-Cerep/P270, accessed June 29, 2020).

62. J. J. Lynch, T. R. Van Vleet, S. W. Mittelstadt, E. A. G. Blomme, Potential functional and pathological side effects related to off-target pharmacological activity, *J. Pharmacol. Toxicol. Methods* **87**, 108–126 (2017).

63. E. Lounkine, M. J. Keiser, S. Whitebread, D. Mikhailov, J. Hamon, J. L. Jenkins, P. Lavan, E. Weber, A. K. Doak, S. Côté, B. K. Shoichet, L. Urban, Large-scale prediction and testing of drug activity on side-effect targets, *Nature* **486**, 361–367 (2012).

64. A. M. Deaton, F. Fan, W. Zhang, P. A. Nguyen, L. D. Ward, P. Nioi, Rationalizing secondary pharmacology screening using human genetic and pharmacological evidence, *Toxicol. Sci.* **167**, 593–603 (2019).

65. S. Bendels, C. Bissantz, B. Fasching, G. Gerebtzoff, W. Guba, M. Kansy, J. Migeon, S. Mohr, J.-U. Peters, F. Tillier, R. Wyler, C. Lerner, C. Kramer, H. Richter, S. Roberts, Safety screening in early drug discovery: an optimized assay panel, *J. Pharmacol. Toxicol. Methods* **99**, 106609 (2019).

66. L. Urban, S. Whitebread, J. Hamon, D. Mikhailov, K. Azzaoui, in *Polypharmacology in Drug Discovery*, (John Wiley & Sons, Ltd, Hoboken, New Jersey, 2012), pp. 15–46.

67. S. Whitebread, B. Dumotier, D. Armstrong, A. Fekete, S. Chen, A. Hartmann, P. Y. Muller, L. Urban, Secondary pharmacology: screening and interpretation of off-target activities – focus on translation, *Drug Discov. Today* **21**, 1232–1242 (2016).

68. M. Awale, J.-L. Reymond, The polypharmacology browser: a web-based multi-fingerprint target prediction tool using ChEMBL bioactivity data, *J. Cheminformatics* **9**, 11 (2017).

69. C. M. Krejsa, D. Horvath, S. L. Rogalski, J. E. Penzotti, B. Mao, F. Barbosa, J. C. Migeon, Predicting ADME properties and side effects: the BioPrint approach, *Curr. Opin. Drug Discov. Devel.* **6**, 470–480 (2003).

70. T. Papoian, H.-J. Chiu, I. Elayan, G. Jagadeesh, I. Khan, A. A. Laniyonu, C. X. Li, M. Saulnier, N. Simpson, B. Yang, Secondary pharmacology data to assess potential off-target activity of new drugs: a regulatory perspective, *Nat. Rev. Drug Discov.* **14**, 294 (2015).

71. C. E. Pollard, M. Skinner, S. E. Lazic, H. M. Prior, K. M. Conlon, J.-P. Valentin, C. Dota, An analysis of the relationship between preclinical and clinical QT interval-related data, *Toxicol. Sci.* **159**, 94–101 (2017).

72. J.-P. Valentin, T. Hammond, Safety and secondary pharmacology: successes, threats, challenges and opportunities, *J. Pharmacol. Toxicol. Methods* **58**, 77–87 (2008).

73. M. Maciejewski, E. Lounkine, S. Whitebread, P. Farmer, W. DuMouchel, B. K. Shoichet, L. Urban, Reverse translation of adverse event reports paves the way for de-risking preclinical off-targets, *eLife* **6**, e25818 (2017).

74. M. Clark, T. Steger-Hartmann, A big data approach to the concordance of the toxicity of pharmaceuticals in animals and humans, *Regul. Toxicol. Pharmacol.* **96**, 94–105 (2018).

75. P. Y. Muller, D. Dambach, B. Gemzik, A. Hartmann, S. Ratcliffe, C. Trendelenburg, L. Urban, Integrated risk assessment of suicidal ideation and behavior in drug development, *Drug Discov. Today* **20**, 1135–1142 (2015).

76. T. Sameshima, T. Yukawa, Y. Hirozane, M. Yoshikawa, T. Katoh, H. Hara, T. Yogo, I. Miyahisa, T. Okuda, M. Miyamoto, R. Naven, Small-scale panel comprising diverse gene family targets to evaluate compound promiscuity, *Chem. Res. Toxicol.* (2019).

77. J. D. Hughes, J. Blagg, D. A. Price, S. Bailey, G. A. DeCrescenzo, R. V. Devraj, E. Ellsworth, Y. M. Fobian, M. E. Gibbs, R. W. Gilles, N. Greene, E. Huang, T. Krieger-Burke, J. Loesel, T. Wager, L. Whiteley, Y. Zhang, Physiochemical drug properties associated with in vivo toxicological outcomes, *Bioorg. Med. Chem. Lett.* **18**, 4872–4875 (2008).

78. D. A. Price, J. Blagg, L. Jones, N. Greene, T. Wager, Physicochemical drug properties associated with in vivo toxicological outcomes: a review, *Expert Opin. Drug Metab. Toxicol.* **5**, 921–931 (2009).

79. K. Azzaoui, J. Hamon, B. Faller, S. Whitebread, E. Jacoby, A. Bender, J. L. Jenkins, L. Urban, Modeling promiscuity based on in vitro safety pharmacology profiling data, *ChemMedChem* **2**, 874–880 (2007).

80. T. M. Monticello, T. W. Jones, D. M. Dambach, D. M. Potter, M. W. Bolt, M. Liu, D. A. Keller, T. K. Hart, V. J. Kadambi, Current nonclinical testing paradigm enables safe entry to first-in-human clinical trials: the IQ consortium nonclinical to clinical translational database, *Toxicol. Appl. Pharmacol.* **334**, 100–109 (2017).

81. M. Clark, Prediction of clinical risks by analysis of preclinical and clinical adverse events, *J. Biomed. Inform.* **54**, 167–173 (2015).

82. R. C. Zink, O. Marchenko, M. Sanchez-Kam, H. Ma, Q. Jiang, Sources of safety data and statistical strategies for design and analysis: clinical trials, *Ther. Innov. Regul. Sci.* **52**, 141–158 (2018).

83. J. Lexchin, Why are there deadly drugs?, *BMC Med.* **13**, 27 (2015).

84. N. J. DeVito, S. Bacon, B. Goldacre, Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: a cohort study, *The Lancet* **395**, 361–369 (2020).

85. D. M. Hartung, D. A. Zarin, J.-M. Guise, M. McDonagh, R. Paynter, M. Helfand, Reporting discrepancies between the ClinicalTrials.gov results database and peer-reviewed publications, *Ann. Intern. Med.* **160**, 477–483 (2014).

86. Home - Electronic Medicines Compendium, (available at https://www.medicines.org.uk/emc/, accessed June 21, 2020).

87. M. Kuhn, I. Letunic, L. J. Jensen, P. Bork, The SIDER database of drugs and side effects, *Nucleic Acids Res.* **44**, D1075–D1079 (2016).

88. G. Candore, K. Juhlin, K. Manlik, B. Thakrar, N. Quarcoo, S. Seabroke, A. Wisniewski, J. Slattery, Comparison of statistical signal detection methods within and across spontaneous reporting databases, *Drug Saf.* **38**, 577–587 (2015).

89. Questions and Answers on FDA's Adverse Event Reporting System (FAERS), (2018) (available at

https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm, accessed June 29, 2020).

90. W. P. Stephenson, M. Hauben, Data mining for signals in spontaneous reporting databases: proceed with caution, *Pharmacoepidemiol. Drug Saf.* **16**, 359–365 (2007).

91. FAERS Quarterly Data Extract Files, (available at https://fis.fda.gov/extensions/FPD-QDE-FAERS/FPD-QDE-FAERS.html, accessed March 12, 2020).

92. Uppsala Monitoring Centre, What is VigiBase?, (available at https://www.who-umc.org/vigibase/vigibase/, accessed June 29, 2020).

93. European Medicines Agency, EudraVigilance, (available at https://www.ema.europa.eu/en/human-regulatory/research-development/pharmacovigilance/eudravigilance, accessed June 29, 2020).

94. R. Harpaz, W. DuMouchel, N. H. Shah, D. Madigan, P. Ryan, C. Friedman, Novel data mining methodologies for adverse drug event discovery and analysis, *Clin. Pharmacol. Ther.* **91**, 1010–1021 (2012).

95. R. B. D'Agostino, Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group, *Stat. Med.* **17**, 2265–2281 (1998).

96. P. C. Austin, A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003, *Stat. Med.* **27**, 2037–2049 (2008).

97. J. Talbot, J. K. Aronson, Eds., *Stephens' Detection and Evaluation of Adverse Drug Reactions: Principles and Practice* (Wiley-Blackwell, Chichester, 6th edition, 2012).

98. N. P. Tatonetti, P. P. Ye, R. Daneshjou, R. B. Altman, Data-driven prediction of drug effects and interactions, *Sci. Transl. Med.* **4**, 125ra31-125ra31 (2012).

99. A. Bate, S. J. W. Evans, Quantitative signal detection using spontaneous ADR reporting, *Pharmacoepidemiol. Drug Saf.* **18**, 427–436 (2009).

100. P. Ghosh, A. Dewanji, Effect of reporting bias in the analysis of spontaneous reporting data, *Pharm. Stat.* **14**, 20–25 (2015).

101. Y. Li, H. Salmasian, S. Vilar, H. Chase, C. Friedman, Y. Wei, A method for controlling complex confounding effects in the detection of adverse drug reactions using electronic health records, *J. Am. Med. Inform. Assoc.* **21**, 308–314 (2014).

102. M. A. Brookhart, S. Schneeweiss, K. J. Rothman, R. J. Glynn, J. Avorn, T. Stürmer, Variable selection for propensity score models, *Am. J. Epidemiol.* **163**, 1149–1156 (2006).

103. P. A. Nguyen, D. A. Born, A. M. Deaton, P. Nioi, L. D. Ward, Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects, *Nat. Commun.* **10**, 1579 (2019).

104. G. Nicola, T. Liu, M. K. Gilson, Public domain databases for medicinal chemistry, *J. Med. Chem.* **55**, 6987–7002 (2012).

105. A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, ChEMBL: a large-scale bioactivity database for drug discovery, *Nucleic Acids Res.* **40**, D1100–D1107 (2012).

106. D. Mendez, A. Gaulton, A. P. Bento, J. Chambers, M. De Veij, E. Félix, M. P. Magariños, J. F. Mosquera, P. Mutowo, M. Nowotka, M. Gordillo-Marañón, F. Hunter, L. Junco, G. Mugumbate, M. Rodriguez-Lopez, F. Atkinson, N. Bosc, C. J. Radoux, A. Segura-Cabrera, A. Hersey, A. R. Leach, ChEMBL: towards direct deposition of bioassay data, *Nucleic Acids Res.* **47**, D930–D940 (2019).

107. N. Bosc, F. Atkinson, E. Felix, A. Gaulton, A. Hersey, A. R. Leach, Large scale comparison of QSAR and conformal prediction methods and their applications in drug discovery, *J. Cheminformatics* **11**, 4 (2019).

108. J. Sun, N. Jeliazkova, V. Chupakhin, J.-F. Golib-Dzib, O. Engkvist, L. Carlsson, J. Wegner, H. Ceulemans, I. Georgiev, V. Jeliazkov, N. Kochev, T. J. Ashby, H. Chen, ExCAPE-DB: an integrated large scale dataset facilitating Big Data analysis in chemogenomics, *J. Cheminformatics* **9**, 17 (2017).

109. D. Szklarczyk, A. Santos, C. von Mering, L. J. Jensen, P. Bork, M. Kuhn, STITCH 5: augmenting protein–chemical interaction networks with tissue and affinity data, *Nucleic Acids Res.* **44**, D380–D384 (2016).

110. D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018, *Nucleic Acids Res.* **46**, D1074–D1082 (2018).

111. R. Huang, M. Xia, S. Sakamuru, J. Zhao, S. A. Shahane, M. Attene-Ramos, T. Zhao, C. P. Austin, A. Simeonov, Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction and mechanism characterization, *Nat. Commun.* **7**, 10425 (2016).

112. S. Lampa, J. Alvarsson, S. Arvidsson Mc Shane, A. Berg, E. Ahlberg, O. Spjuth, Predicting off-target binding profiles with confidence using Conformal Prediction, *Front. Pharmacol.* **9** (2018).

113. M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin, B. K. Shoichet, Relating protein pharmacology by ligand chemistry, *Nat. Biotechnol.* **25**, 197–206 (2007).

114. L. H. Mervin, K. C. Bulusu, L. Kalash, A. M. Afzal, F. Svensson, M. A. Firth, I. Barrett, O. Engkvist, A. Bender, Orthologue chemical space and its influence on target prediction, *Bioinformatics* **34**, 72–79 (2018).

115. D. Rogers, M. Hahn, Extended-Connectivity Fingerprints, *J. Chem. Inf. Model.* **50**, 742–754 (2010).

116. Prediction IncluDinG INactivity (PIDGIN) Version 3 — PIDGINv3 v0.1beta documentation, (available at https://pidginv3.readthedocs.io/en/latest/, accessed September 5, 2019).

117. D. Sydow, L. Burggraaff, A. Szengel, H. W. T. van Vlijmen, A. P. IJzerman, G. J. P. van Westen, A. Volkamer, Advances and challenges in computational target prediction, *J. Chem. Inf. Model.* **59**, 1728–1742 (2019).

118. N. Aniceto, A. A. Freitas, A. Bender, T. Ghafourian, A novel applicability domain technique for mapping predictive reliability across the chemical space of a QSAR: reliability-density neighbourhood, *J. Cheminformatics* **8**, 69 (2016).

119. M. Kuhn, M. Campillos, I. Letunic, L. J. Jensen, P. Bork, A side effect resource to capture phenotypic effects of drugs, *Mol. Syst. Biol.* **6**, 343 (2010).

120. SIDER Side Effect Resource, (available at http://sideeffects.embl.de/download/, accessed March 23, 2020).

121. J. M. Banda, L. Evans, R. S. Vanguri, N. P. Tatonetti, P. B. Ryan, N. H. Shah, A curated and standardized adverse drug event resource to accelerate drug safety research, *Sci. Data* **3**, 160026 (2016).

122. K. Kreimeyer, D. Menschik, S. Winiecki, W. Paul, F. Barash, E. J. Woo, M. Alimchandani, D. Arya, C. Zinderman, R. Forshee, T. Botsis, Using probabilistic record linkage of structured and unstructured data to identify duplicate cases in spontaneous adverse event reporting systems, *Drug Saf.* **40** (2017).

123. I. Grigoriev, W. zu Castell, P. Tsvetkov, A. V. Antonov, AERS spider: an online interactive tool to mine statistical associations in Adverse Event Reporting System, *Pharmacoepidemiol. Drug Saf.* **23**, 795–801 (2014).

124. C. K. Wong, S. S. Ho, B. Saini, D. E. Hibbs, R. A. Fois, Standardisation of the FAERS database: a systematic approach to manually recoding drug name variants, *Pharmacoepidemiol. Drug Saf.* **24**, 731–737 (2015).

125. E. P. van Puijenbroek, A. Bate, H. G. M. Leufkens, M. Lindquist, R. Orre, A. C. G. Egberts, A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions, *Pharmacoepidemiol. Drug Saf.* **11**, 3–10 (2002).

126. B. H. Ch. Stricker, J. G. P. Tijssen, Serum sickness-like reactions to cefaclor, *J. Clin. Epidemiol.* **45**, 1177–1184 (1992).

127. A. Bate, M. Lindquist, I. R. Edwards, S. Olsson, R. Orre, A. Lansner, R. M. De Freitas, A Bayesian neural network method for adverse drug reaction signal generation, *Eur. J. Clin. Pharmacol.* **54**, 315–321 (1998).

128. O. Caster, G. N. Norén, D. Madigan, A. Bate, Large-scale regression-based pattern discovery: the example of screening the WHO global drug safety database, *Stat. Anal. Data Min.* **3**, 197–208 (2010).

129. W. DuMouchel, Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system, *Am. Stat.* **53**, 177–190 (1999).

130. Yan C., Jeff J. G., Michael S., Xiaodong L., Charles R. B., Nick C. P., Comparison of sensitivity and timing of early signal detection of four frequently used signal detection methods, *Pharm. Med.* **22**, 359–365 (2008).

131. L. Miguel de Almeida Vieira Lima, N. Goncalo Sales Craveiro Nunes, P. Goncalo Pires da Silva Dias, F. Jorge Batel Marques, Implemented data mining and signal management systems on spontaneous reporting systems' databases and their availability to the scientific community - A systematic review, *Curr. Drug Saf.* **7**, 170–175 (2012).

132. J. Hopstadius, G. N. Norén, A. Bate, I. R. Edwards, Impact of stratification on adverse drug reaction surveillance, *Drug Saf.* **31**, 1035–1048 (2008).

133. U.S. Food and Drug Administration, Data mining at FDA - White Paper, (2018) (available at https://www.fda.gov/ScienceResearch/DataMiningatFDA/ucm446239.htm, accessed June 29, 2020).

134. S. Seabroke, G. Candore, K. Juhlin, N. Quarcoo, A. Wisniewski, R. Arani, J. Painter, P. Tregunno, G. N. Norén, J. Slattery, Performance of stratified and subgrouped disproportionality analyses in spontaneous databases, *Drug Saf.* **39**, 355–364 (2016).

135. R. Solomon, W. Dumouchel, Contrast media and nephropathy: findings from systematic analysis and Food and Drug Administration reports of adverse effects, *Invest. Radiol.* **41**, 651–660 (2006).

136. W. M. Holmes, *Using Propensity Scores in Quasi-experimental Designs* (Sage Publications, Los Angeles, California, 2014).

137. P. C. Austin, An introduction to propensity score methods for reducing the effects of confounding in observational studies, *Multivar. Behav. Res.* **46**, 399–424 (2011).

138. S. Deb, P. C. Austin, J. V. Tu, D. T. Ko, C. D. Mazer, A. Kiss, S. E. Fremes, A review of propensity-score methods and their use in cardiovascular research, *Can. J. Cardiol.* **32**, 259–265 (2016).

139. G. Heinze, P. Jüni, An overview of the objectives of and the approaches to propensity score analyses, *Eur. Heart J.* **32**, 1704–1708 (2011).

140. S. Schneeweiss, J. A. Rassen, R. J. Glynn, J. Avorn, H. Mogun, M. A. Brookhart, High-dimensional propensity score adjustment in studies of treatment effects using health care claims data, *Epidemiology* **20**, 512–522 (2009).

141. M. G. Coulthard, Quantifying how tests reduce diagnostic uncertainty, *Arch. Dis. Child.* **92**, 404–408 (2007).

142. K. Johnson, M. Kuhn, *Applied Predictive Modeling* (Springer, New York, 2013).

143. D. G. Altman, J. M. Bland, Diagnostic tests 2: predictive values, *BMJ* **309**, 102 (1994).

144. D. A. Grimes, K. F. Schulz, Uses and abuses of screening tests, *The Lancet* **359**, 881–884 (2002).

145. D. A. Grimes, K. F. Schulz, Refining clinical diagnosis with likelihood ratios, *The Lancet* **365**, 1500–1505 (2005).

146. P. McClure, Likelihood ratios: determining the usefulness of diagnostic tests, *J. Hand Ther.* **14**, 304–305 (2001).

147. P. F. W. Chien, K. S. Khan, Evaluation of a clinical test. II: Assessment of validity, *Br. J. Obstet. Gynaecol.* **108**, 568–572 (2001).

148. A. F. Fliri, W. T. Loging, P. F. Thadeio, R. A. Volkmann, Analysis of drug-induced effect patterns to link structure and side effects of medicines, *Nat. Chem. Biol.* **1**, 389–397 (2005).

149. A. Bender, J. Scheiber, M. Glick, J. W. Davies, K. Azzaoui, J. Hamon, L. Urban, S. Whitebread, J. L. Jenkins, Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure, *ChemMedChem* **2**, 861–873 (2007).

150. M. Kuhn, M. A. Banchaabouchi, M. Campillos, L. J. Jensen, C. Gross, A.-C. Gavin, P. Bork, Systematic identification of proteins that elicit drug side effects, *Mol. Syst. Biol.* **9**, 663 (2013).

151. M. Duran-Frigola, P. Aloy, Analysis of chemical and biological features yields mechanistic insights into drug side effects, *Chem. Biol.* **20**, 594–603 (2013).

152. W. S. Redfern, L. Carlsson, A. S. Davis, W. G. Lynch, I. MacKenzie, S. Palethorpe, P. K. S. Siegl, I. Strang, A. T. Sullivan, R. Wallis, A. J. Camm, T. G. Hammond, Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for

a broad range of drugs: evidence for a provisional safety margin in drug development, *Cardiovasc. Res.* **58**, 32–45 (2003).

153. N. S. Sipes, J. F. Wambaugh, R. Pearce, S. S. Auerbach, B. A. Wetmore, J.-H. Hsieh, A. J. Shapiro, D. Svoboda, M. J. DeVito, S. S. Ferguson, An intuitive approach for predicting potential human health risk with the Tox21 10k library, *Environ. Sci. Technol.* **51**, 10786–10796 (2017).

154. F. Svensson, A. Zoufir, S. Mahmoud, A. M. Afzal, I. Smit, K. A. Giblin, P. J. Clements, J. T. Mettetal, A. Pointon, J. S. Harvey, N. Greene, R. V. Williams, A. Bender, Information-derived mechanistic hypotheses for structural cardiotoxicity, *Chem. Res. Toxicol.* **31**, 1119–1127 (2018).

155. L. Wang, G. Jiang, D. Li, H. Liu, Standardizing adverse drug event reporting data, *J. Biomed. Semant.* **5**, 36 (2014).

156. E. W. Boyer, M. Shannon, The serotonin syndrome, *N. Engl. J. Med.* **352**, 1112–1120 (2005).

157. V. L. Culbertson, S. E. Rahman, G. C. Bosen, M. L. Caylor, M. M. Echevarria, D. Xu, Implications of off-target serotoninergic drug activity - An analysis of serotonin syndrome reports using a systematic bioinformatics approach, *Pharmacotherapy* (2018).

158. R. Racz, T. G. Soldatos, D. Jackson, K. Burkhart, Association between serotonin syndrome and second-generation antipsychotics via pharmacological target-adverse event analysis, *Clin. Transl. Sci.* **11**, 322–329 (2018).

159. R. Ietswaart, S. Arat, A. X. Chen, S. Farahmand, B. Kim, W. DuMouchel, D. Armstrong, A. Fekete, J. J. Sutherland, L. Urban, Machine learning guided association of adverse drug reactions with in vitro target-based pharmacology, *EBioMedicine* **57** (2020), doi:10.1016/j.ebiom.2020.102837.

160. A. F. Z. Wisniewski, A. Bate, C. Bousquet, A. Brueckner, G. Candore, K. Juhlin, M. A. Macia-Martinez, K. Manlik, N. Quarcoo, S. Seabroke, J. Slattery, H. Southworth, B. Thakrar, P. Tregunno, L. V. Holle, M. Kayser, G. N. Norén, Good signal detection practices: evidence from IMI PROTECT, *Drug Saf.* **39**, 469–490 (2016).

161. RxClass Overview, (available at https://rxnav.nlm.nih.gov/RxClassIntro.html, accessed June 29, 2020).

162. Documentation scikit-learn: machine learning in Python — scikit-learn 0.21.3 documentation, (available at https://scikit-learn.org/stable/documentation.html, accessed September 4, 2019).

163. Statistical functions (scipy.stats) — SciPy Reference Guide, (available at https://docs.scipy.org/doc/scipy/reference/stats.html, accessed June 29, 2020).

164. RxNorm, (available at https://www.nlm.nih.gov/research/umls/rxnorm/, accessed June 29, 2020).

165. M. Kuhn, *SIDER – Side Effect Resource* (2019; https://github.com/mkuhn/sider).

166. J. Chambers, M. Davies, A. Gaulton, A. Hersey, S. Velankar, R. Petryszak, J. Hastings, L. Bellis, S. McGlinchey, J. P. Overington, UniChem: a unified chemical structure cross-referencing and identifier tracking system, *J. Cheminformatics* **5**, 3 (2013).

167. A. Gaulton, A. Hersey, M. Nowotka, A. P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrián-Uhalte, M. Davies, N. Dedman, A. Karlsson, M. P. Magariños, J. P. Overington, G. Papadatos, I. Smit, A. R. Leach, The ChEMBL database in 2017, *Nucleic Acids Res.* **45**, D945–D954 (2017).

168. DailyMed - PROCHLORPERAZINE MALEATE tablet, film coated, (available at https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=6b0e958b-fde0-74ea-b196-1e62f90b5bbe, accessed June 23, 2020).

169. DailyMed - ONDANSETRON- ondansetron hydrochloride injection, (available at https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=e0050959-c14c-41b6-9a92-fadc5f6feff3, accessed June 23, 2020).

170. K. F. Huybrechts, B. T. Bateman, S. Hernández-Díaz, Use of real-world evidence from healthcare utilization data to evaluate drug safety during pregnancy, *Pharmacoepidemiol. Drug Saf.* **28**, 906–922 (2019).

171. J. Scaffidi, B. W. Mol, J. A. Keelan, The pregnant women as a drug orphan: a global survey of registered clinical trials of pharmacological interventions in pregnancy, *Br. J. Obstet. Gynaecol.* **124**, 132–140 (2017).

172. A. R. Patrick, S. Schneeweiss, M. A. Brookhart, R. J. Glynn, K. J. Rothman, J. Avorn, T. Stürmer, The implications of propensity score variable selection strategies in pharmacoepidemiology: an empirical illustration, *Pharmacoepidemiol. Drug Saf.* **20**, 551–559 (2011).

173. D. Westreich, S. R. Cole, M. J. Funk, M. A. Brookhart, T. Stürmer, The role of the c-statistic in variable selection for propensity score models, *Pharmacoepidemiol. Drug Saf.* **20**, 317–320 (2011).

174. A. Z. Fu, L. Li, Thinking of having a higher predictive power for your first-stage model in propensity score analysis? Think again, *Health Serv. Outcomes Res. Methodol.* **8**, 115–117 (2008).

175. F. Atkinson, *Molecular standardisation tool* (2019; https://github.com/flatkinson/standardiser).

176. M. Schulz, S. Iwersen-Bergmann, H. Andresen, A. Schmoldt, Therapeutic and toxic blood concentrations of nearly 1,000 drugs and other xenobiotics, *Crit. Care* **16**, R136 (2012).

177. F. Zhang, J. Xue, J. Shao, L. Jia, Compilation of 222 drugs' plasma protein binding data and guidance for study designs, *Drug Discov. Today* **17**, 475–485 (2012).

178. F. Lombardo, G. Berellini, R. S. Obach, Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 1352 drug compounds, *Drug Metab. Dispos.* **46**, 1466–1477 (2018).

179. C. Kramer, T. Kalliokoski, P. Gedeck, A. Vulpetti, The experimental uncertainty of heterogeneous public Ki data, *J. Med. Chem.* **55**, 5165–5173 (2012).

180. GSK Published Kinase Inhibitor Set - Nanosyn kinase panel, Document Report Card, (available at https://www.ebi.ac.uk/chembl/document_report_card/CHEMBL1961873/, accessed May 4, 2020).

181. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

182. Statsmodels, (available at https://www.statsmodels.org/stable/index.html, accessed April 13, 2020).

183. P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, M. A. Musen, BioPortal: Enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications, *Nucleic Acids Res.* **39**, W541-545 (2011).

184. J. Mischlinger, E. Schernhammer, A common trap of diagnostic tests, *Wien. Klin. Wochenschr.* **129**, 583–584 (2017).

185. I. Grenet, J. P. Comet, F. Schorsch, N. Ryan, J. Wichard, D. Rouquié, Chemical in vitro bioactivity profiles are not informative about the long-term in vivo endocrine mediated toxicity, *Comput. Toxicol.* **12**, 100098 (2019).

186. S. Y. Mahmoud, F. Svensson, A. Zoufir, D. Módos, A. M. Afzal, A. Bender, Understanding conditional associations between ToxCast in vitro readouts and the hepatotoxicity of compounds using rule-based methods, *Chem. Res. Toxicol.* (2019).

187. M. Chen, V. Vijay, Q. Shi, Z. Liu, H. Fang, W. Tong, FDA-approved drug labeling for the study of drug-induced liver injury, *Drug Discov. Today* **16**, 697–703 (2011).

188. C. Diez-Fernandez, V. Rüfenacht, S. Santra, A. M. Lund, R. Santer, M. Lindner, T. Tangeraas, C. Unsinn, P. de Lonlay, A. Burlina, C. D. M. van Karnebeek, J. Häberle, Defective hepatic bicarbonate production due to carbonic anhydrase VA deficiency leads to early-onset life-threatening metabolic crisis, *Genet. Med.* **18**, 991–1000 (2016).

189. A. Q. T. Pham, L. H. R. Xu, O. W. Moe, Drug-induced metabolic acidosis, *F1000Research* **4**, 1460 (2015).

190. DailyMed - ACETAZOLAMIDE tablet, (available at https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=ceb5cac3-bbfb-42fa-8b29-c0ef5ed0913c, accessed June 29, 2020).

191. CA5B protein expression summary - The Human Protein Atlas, (available at https://www.proteinatlas.org/ENSG00000169239-CA5B, accessed April 6, 2020).

192. A. Ramachandran, R. G. J. Visschers, L. Duan, J. Y. Akakpo, H. Jaeschke, Mitochondrial dysfunction as a mechanism of drug-induced hepatotoxicity: current understanding and future perspectives, *J. Clin. Transl. Res.* **4**, 75–100 (2018).

193. F. Fang, H. Sun, Z. Wang, M. Ren, J. R. Calabrese, K. Gao, Antipsychotic drug-induced somnolence: incidence, mechanisms, and management, *CNS Drugs* **30**, 845–867 (2016).

194. A. Paquette, P. Gou, C. Tannenbaum, Systematic review and meta-analysis: do clinical trials testing antimuscarinic agents for overactive bladder adequately measure central nervous system adverse events?, *J. Am. Geriatr. Soc.* **59**, 1332–1339 (2011).

195. Tolterodine Compound Report Card, (available at https://www.ebi.ac.uk/chembl/compound_report_card/CHEMBL1382/, accessed February 11, 2020).

196. Oxybutynin Compound Report Card, (available at https://www.ebi.ac.uk/chembl/compound_report_card/CHEMBL1231/, accessed February 11, 2020).

197. J.-L. Vincent, D. De Backer, Circulatory shock, *N. Engl. J. Med.* **369**, 1726–1734 (2013).

198. H. Charbonneau, M. Buléon, B. Richard, N. Mayeur, Icatibant as an early rescue therapy in hypovolemic shock with converting enzyme inhibitor treatment, *Crit. Care* **21** (2017).

199. J. Myslinski, A. Heiser, A. Kinney, Hypovolemic shock caused by angiotensin-converting enzyme inhibitor-induced visceral angioedema: A case series and a simple method to diagnose this complication in the emergency department, *J. Emerg. Med.* **54**, 375–379 (2018).

200. T. Günther, P. Dasgupta, A. Mann, E. Miess, A. Kliewer, S. Fritzwanker, R. Steinborn, S. Schulz, Targeting multiple opioid receptors – improved analgesics with reduced side effects?, *Br. J. Pharmacol.* **175**, 2857–2868 (2018).

201. CA5A protein expression summary - The Human Protein Atlas, (available at https://www.proteinatlas.org/ENSG00000174990-CA5A, accessed June 11, 2020).

202. G. M. Kennedy, S. D. Lhatoo, CNS adverse events associated with antiepileptic drugs, *CNS Drugs* **22**, 739–760 (2008).

203. D. Feurstein, K. Stemmer, J. Kleinteich, T. Speicher, D. R. Dietrich, Microcystin congener- and concentration-dependent induction of murine neuron apoptosis and neurite degeneration, *Toxicol. Sci.* **124**, 424–431 (2011).

204. R. Velamoor, Neuroleptic malignant syndrome: a neuro-psychiatric emergency: recognition, prevention, and management, *Asian J. Psychiatry* **29**, 106–109 (2017).

205. B. H. Mullish, H. R. Williams, Clostridium difficile infection and antibiotic-associated diarrhoea, *Clin. Med.* **18**, 237–241 (2018).

206. N. P. Tatonetti, Data-driven detection, prediction, and validation of drug-drug interactions, thesis, Stanford University (2012).

207. T. E. H. Allen, S. Liggi, J. M. Goodman, S. Gutsell, P. J. Russell, Using molecular initiating events to generate 2D structure–activity relationships for toxicity screening, *Chem. Res. Toxicol.* **29**, 1611–1627 (2016).

208. A. Mayr, G. Klambauer, T. Unterthiner, S. Hochreiter, DeepTox: Toxicity prediction using Deep Learning, *Front. Environ. Sci.* **3** (2016), doi:10.3389/fenvs.2015.00080.

209. M. Fernandez, F. Ban, G. Woo, M. Hsing, T. Yamazaki, E. LeBlanc, P. S. Rennie, W. J. Welch, A. Cherkasov, Toxic Colors: the use of Deep Learning for predicting toxicity of compounds merely from their graphic images, *J. Chem. Inf. Model.* **58**, 1533–1543 (2018).

210. T. Cho, J. Uetrecht, How reactive metabolites induce an immune response that sometimes leads to an idiosyncratic drug reaction, *Chem. Res. Toxicol.* **30**, 295–314 (2017).

211. T. I. Oprea, C. G. Bologa, S. Brunak, A. Campbell, G. N. Gan, A. Gaulton, S. M. Gomez, R. Guha, A. Hersey, J. Holmes, A. Jadhav, L. J. Jensen, G. L. Johnson, A. Karlson, A. R. Leach, A. Ma'ayan, A. Malovannaya, S. Mani, S. L. Mathias, M. T. McManus, T. F. Meehan, C. von Mering, D. Muthas, D.-T. Nguyen, J. P. Overington, G. Papadatos, J. Qin, C. Reich, B. L. Roth, S. C. Schürer, A. Simeonov, L. A. Sklar, N. Southall, S. Tomita, I. Tudose, O. Ursu, D. Vidović, A. Waller, D. Westergaard, J. J. Yang, G. Zahoránszky-Köhalmi, Unexplored therapeutic opportunities in the human genome, *Nat. Rev. Drug Discov.* **17**, 317–332 (2018).

212. F. Sanz, F. Pognan, T. Steger-Hartmann, C. Díaz, eTOX, M. Cases, M. Pastor, P. Marc, J. Wichard, K. Briggs, D. K. Watson, T. Kleinöder, C. Yang, A. Amberg, M. Beaumont, A. J. Brookes, S. Brunak, M. T. D. Cronin, G. F. Ecker, S. Escher, N. Greene, A. Guzmán, A. Hersey, P. Jacques, L. Lammens, J. Mestres, W. Muster, H. Northeved, M. Pinches, J. Saiz, N. Sajot, A. Valencia, J. van der Lei, N. P. E. Vermeulen, E. Vock, G. Wolber, I. Zamora, Legacy data sharing to improve drug safety assessment: the eTOX project, *Nat. Rev. Drug Discov.* **16**, 811–812 (2017).

213. T. Hanser, L. Johnston, J. Marchaland, J. Plante, R. van Deursen, S. Werner, R. Williams, From Private Data to Shared Knowledge, poster, *Lhasa Ltd.* (2020) (available at https://www.lhasalimited.org/publications/from-private-data-to-shared-knowledge/6686, accessed June 18, 2020).

214. T. Hanser, F. P. Steinmetz, J. Plante, F. Rippmann, M. Krier, Avoiding hERG-liability in drug design via synergetic combinations of different (Q)SAR methodologies and data sources: a case study in an industrial setting, *J. Cheminformatics* **11**, 9 (2019).

215. R. Huang, M. Xia, M.-H. Cho, S. Sakamuru, P. Shinn, K. A. Houck, D. J. Dix, R. S. Judson, K. L. Witt, R. J. Kavlock, R. R. Tice, C. P. Austin, Chemical genomics profiling of environmental chemical modulation of human nuclear receptors, *Environ. Health Perspect.* **119**, 1142–1148 (2011).

216. J.-H. Hsieh, A. Sedykh, R. Huang, M. Xia, R. R. Tice, A data analysis pipeline accounting for artifacts in Tox21 quantitative high-throughput screening assays, *J. Biomol. Screen.* **20**, 887–897 (2015).

217. D. L. Filer, P. Kothiya, R. W. Setzer, R. S. Judson, M. T. Martin, tcpl: the ToxCast pipeline for high-throughput screening data, *Bioinformatics* **33**, 618–620 (2017).

218. M. T. Martin, T. B. Knudsen, D. M. Reif, K. A. Houck, R. S. Judson, R. J. Kavlock, D. J. Dix, Predictive model of rat reproductive toxicity from ToxCast high throughput screening, *Biol. Reprod.* **85**, 327–339 (2011).

219. J. Liu, K. Mansouri, R. S. Judson, M. T. Martin, H. Hong, M. Chen, X. Xu, R. S. Thomas, I. Shah, Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure, *Chem. Res. Toxicol.* **28**, 738–751 (2015).

220. R. G. Pearce, R. Woodrow Setzer, C. L. Strope, N. S. Sipes, J. F. Wambaugh, httk: R package for high-throughput toxicokinetics, *J. Stat. Softw.* **79** (2017).

221. G. S. Honda, R. G. Pearce, L. L. Pham, R. W. Setzer, B. A. Wetmore, N. S. Sipes, J. Gilbert, B. Franz, R. S. Thomas, J. F. Wambaugh, Using the concordance of in vitro and in vivo data to evaluate extrapolation assumptions, *PLOS ONE* **14**, e0217564 (2019).

222. N. Spinu, M. T. D. Cronin, S. J. Enoch, J. C. Madden, A. P. Worth, Quantitative adverse outcome pathway (qAOP) models for toxicity prediction, *Arch. Toxicol.* (2020).

223. P. Bloomingdale, C. Housand, J. F. Apgar, B. L. Millard, D. E. Mager, J. M. Burke, D. K. Shah, Quantitative systems toxicology, *Curr. Opin. Toxicol.* **4**, 79–87 (2017).

7 Appendices

Appendix 1

Details for the target-AE associations with the highest Positive Predictive Values (PPVs) in the FAERS dataset based on the adjusted activity calls that integrate the unbound drug plasma concentration with measured and predicted *in vitro* bioactivities. LR=Likelihood Ratio. AE=adverse event. SOC=System Organ Class. Only the 50 most highly ranked are shown but the full data is available in the University of Cambridge repository (https://doi.org/10.17863/CAM.53868).

Target	AE	Drugs (n)	Drugs with AE (n)	Fraction of AE-associated drugs that are active	Fraction of other drugs that are active	LR	Corrected p-value	PPV	Fraction of drugs with AE	Specificity	SOC
Cyclooxygenase-1	DUODENAL ULCER	455	27	0.3333	0.0047	71.3	1.1e-07	0.82	0.0593	1.00	Gastrointestinal
Cyclooxygenase-2	DUODENAL ULCER	625	31	0.2581	0.0067	38.3	5.9e-06	0.67	0.0496	0.99	Gastrointestinal
Angiotensin-converting enzyme	HYPOVOLAEMIC SHOCK	302	10	0.3000	0.0068	43.8	4.9e-02	0.60	0.0331	0.99	Vascular

Angiotensin-converting enzyme	PANCREATITIS NECROTISING	302	9	0.3333	0.0068	48.8	4.9e-02	0.60	0.0298	0.99	Gastrointestinal
Angiotensin-converting enzyme	LIP OEDEMA	302	6	0.5000	0.0068	74.0	2.5e-02	0.60	0.0199	0.99	Gastrointestinal
α-1a adrenergic receptor	DIABETIC NEPHROPATHY	375	19	0.3158	0.0140	22.5	1.5e-03	0.55	0.0507	0.99	Renal and urinary
Serotonin 7 (5-HT7) receptor	DIABETIC NEPHROPATHY	325	19	0.3158	0.0229	13.8	6.7e-03	0.46	0.0585	0.98	Renal and urinary
α-1b adrenergic receptor	DIABETIC NEPHROPATHY	404	18	0.2778	0.0155	17.9	1.8e-02	0.45	0.0446	0.98	Renal and urinary
Dopamine D4 receptor	DIABETIC NEPHROPATHY	640	27	0.1481	0.0114	13.0	4.5e-02	0.36	0.0422	0.99	Renal and urinary
hERG	TORSADE DE POINTES	475	40	0.1750	0.0184	9.5	4.6e-02	0.47	0.0842	0.98	Cardiac
α-1a adrenergic receptor	DIABETIC FOOT	375	13	0.3846	0.0166	23.2	2.6e-03	0.45	0.0347	0.98	Skin and subcutaneous tissue

Serotonin 7 (5-HT7) receptor	DIABETIC FOOT	325	13	0.3846	0.0256	15.0	7.5e-03	0.38	0.0400	0.97	Skin and subcutaneous tissue
α-1b adrenergic receptor	DIABETIC FOOT	404	13	0.3077	0.0179	17.2	2.5e-02	0.36	0.0322	0.98	Skin and subcutaneous tissue
α-1b adrenergic receptor	ORTHOSTATIC HYPOTENSION	404	21	0.2381	0.0157	15.2	2.0e-02	0.45	0.0520	0.98	Vascular
α-1a adrenergic receptor	ORTHOSTATIC HYPOTENSION	375	18	0.2222	0.0196	11.3	4.5e-02	0.36	0.0480	0.98	Vascular
δ opioid receptor	RESPIRATORY DEPRESSION	485	19	0.2632	0.0129	20.4	1.9e-02	0.45	0.0392	0.99	Respiratory, thoracic and mediastinal
κ opioid receptor	RESPIRATORY DEPRESSION	472	19	0.2632	0.0155	17.0	3.2e-02	0.42	0.0403	0.98	Respiratory, thoracic and mediastinal
Cyclooxygenase-1	UPPER GASTROINTESTINAL HAEMORRHAGE	455	15	0.3333	0.0136	24.4	3.9e-03	0.45	0.0330	0.99	Gastrointestinal

Cyclooxygenase-1	TUBULOINTERSTITIAL NEPHRITIS	455	29	0.1724	0.0141	12.2	3.8e-02	0.45	0.0637	0.99	Renal and urinary
Dopamine D5 receptor	FEEDING DISORDER NEONATAL	276	5	0.6000	0.0148	40.6	3.7e-02	0.43	0.0181	0.99	Psychiatric
Serotonin 7 (5-HT7) receptor	DIABETIC COMPLICATION	325	12	0.4167	0.0256	16.3	6.7e-03	0.38	0.0369	0.97	Metabolism and nutrition
α-1a adrenergic receptor	DIABETIC COMPLICATION	375	11	0.4545	0.0165	27.6	1.5e-03	0.36	0.0293	0.98	Metabolism and nutrition
α-1b adrenergic receptor	DIABETIC COMPLICATION	404	10	0.4000	0.0178	22.5	1.8e-02	0.36	0.0248	0.98	Metabolism and nutrition
Dopamine D4 receptor	DIABETIC COMPLICATION	640	20	0.2000	0.0113	17.7	2.9e-02	0.36	0.0312	0.99	Metabolism and nutrition
Serotonin 7 (5-HT7) receptor	HYPERGLYCAEMIC HYPEROSMOLAR NONKETOTIC SYNDROME	325	7	0.7143	0.0252	28.4	7.9e-04	0.38	0.0215	0.97	Metabolism and nutrition
Serotonin 7 (5-HT7) receptor	INSULIN-REQUIRING TYPE 2 DIABETES MELLITUS	325	10	0.5000	0.0254	19.7	4.4e-03	0.38	0.0308	0.97	Metabolism and nutrition

α-1b adrenergic receptor	INSULIN-REQUIRING TYPE 2 DIABETES MELLITUS	404	12	0.3333	0.0179	18.7	2.0e-02	0.36	0.0297	0.98	Metabolism and nutrition
Dopamine D4 receptor	INSULIN-REQUIRING TYPE 2 DIABETES MELLITUS	640	13	0.3077	0.0112	27.6	1.6e-02	0.36	0.0203	0.99	Metabolism and nutrition
α-1a adrenergic receptor	INSULIN-REQUIRING TYPE 2 DIABETES MELLITUS	375	10	0.4000	0.0192	20.9	9.9e-03	0.36	0.0267	0.98	Metabolism and nutrition
Serotonin 7 (5-HT7) receptor	DIABETIC RETINOPATHY	325	16	0.3125	0.0259	12.1	1.9e-02	0.38	0.0492	0.97	Eye
α-1a adrenergic receptor	DIABETIC RETINOPATHY	375	12	0.3333	0.0193	17.3	1.6e-02	0.36	0.0320	0.98	Eye
α-2c adrenergic receptor	EXPOSURE VIA INGESTION	545	26	0.3077	0.0250	12.3	1.1e-03	0.38	0.0477	0.97	Injury, poisoning and procedural complications
Serotonin 2c (5-HT2c) receptor	EXPOSURE VIA INGESTION	509	27	0.3704	0.0353	10.5	2.0e-04	0.37	0.0530	0.96	Injury, poisoning and

											procedural complications
Carbonic anhydrase IV	HYPERAMMONAEMIC ENCEPHALOPATHY	310	5	0.6000	0.0164	36.6	3.3e-02	0.38	0.0161	0.98	Nervous system
α-1b adrenergic receptor	TYPE 1 DIABETES MELLITUS	404	14	0.2857	0.0179	15.9	3.0e-02	0.36	0.0347	0.98	Metabolism and nutrition
α-1a adrenergic receptor	TYPE 1 DIABETES MELLITUS	375	15	0.2667	0.0194	13.7	3.0e-02	0.36	0.0400	0.98	Metabolism and nutrition
Dopamine D4 receptor	TYPE 2 DIABETES MELLITUS	640	24	0.1667	0.0114	14.7	3.3e-02	0.36	0.0375	0.99	Metabolism and nutrition
α-1a adrenergic receptor	TYPE 2 DIABETES MELLITUS	375	16	0.2500	0.0195	12.8	3.3e-02	0.36	0.0427	0.98	Metabolism and nutrition
Dopamine D4 receptor	NEUROLEPTIC MALIGNANT SYNDROME	640	26	0.1538	0.0114	13.5	4.2e-02	0.36	0.0406	0.99	Nervous system
Dopamine D4 receptor	GLYCOSURIA	640	15	0.2667	0.0112	23.8	1.7e-02	0.36	0.0234	0.99	Renal and urinary

α-1b adrenergic receptor	GLYCOSURIA	404	12	0.3333	0.0179	18.7	2.0e-02	0.36	0.0297	0.98	Renal and urinary
α-1a adrenergic receptor	GLYCOSURIA	375	10	0.4000	0.0192	20.9	9.9e-03	0.36	0.0267	0.98	Renal and urinary
Cyclooxygenase-1	GASTRIC ULCER PERFORATION	455	12	0.3333	0.0158	21.1	2.1e-02	0.36	0.0264	0.98	Gastrointestinal
Cyclooxygenase-1	GASTRITIS EROSIVE	455	14	0.2857	0.0159	18.0	2.8e-02	0.36	0.0308	0.98	Gastrointestinal
α-1a adrenergic receptor	CEREBROVASCULAR DISORDER	375	11	0.3636	0.0192	18.9	1.2e-02	0.36	0.0293	0.98	Nervous system
Dopamine D4 receptor	CEREBROVASCULAR DISORDER	640	21	0.1905	0.0113	16.8	3.0e-02	0.36	0.0328	0.99	Nervous system
α-1a adrenergic receptor	DIABETIC GASTROPARESIS	375	11	0.3636	0.0192	18.9	1.2e-02	0.36	0.0293	0.98	Gastrointestinal
α-1a adrenergic receptor	GESTATIONAL DIABETES	375	10	0.4000	0.0192	20.9	9.9e-03	0.36	0.0267	0.98	Pregnancy, puerperium and perinatal conditions

Microtubule-associated protein τ	LIVER INJURY	131	11	0.7273	0.1167	6.2	3.2e-02	0.36	0.0840	0.88	Hepatobiliary

Appendix 2

Details for the target-AE associations with the highest Positive Predictive Values (PPVs) in the SIDER dataset based on the adjusted activity calls that integrate the unbound drug plasma concentration with measured and predicted *in vitro* bioactivities. LR=Likelihood Ratio. AE=adverse event. SOC=System Organ Class. Only the 50 most highly ranked are shown but the full data is available in the University of Cambridge repository (https://doi.org/10.17863/CAM.53868).

Target	AE	Drugs (n)	Drugs with AE (n)	Fraction of AE-associated drugs that are active	Fraction of other drugs that are active	LR	Corrected p-value	PPV	Fraction of drugs with AE	Specificity	SOC
Carbonic anhydrase 5B	DECREASED APPETITE	191	67	0.0896	0.0000	inf	1.8e-02	1.00	0.3508	1.00	Metabolism and nutrition
Carbonic anhydrase 9	DECREASED APPETITE	326	78	0.0897	0.0081	11.1	3.1e-02	0.78	0.2393	0.99	Metabolism and nutrition
Carbonic anhydrase 1	DECREASED APPETITE	323	84	0.0833	0.0084	10.0	4.6e-02	0.78	0.2601	0.99	Metabolism and nutrition
Carbonic anhydrase 12	DECREASED APPETITE	294	75	0.1067	0.0137	7.8	2.9e-02	0.73	0.2551	0.99	Metabolism and nutrition
Carbonic anhydrase 5B	PARAESTHESIA	191	80	0.0750	0.0000	inf	3.3e-02	1.00	0.4188	1.00	Nervous system
Carbonic anhydrase 5A	PARAESTHESIA	216	85	0.0706	0.0000	inf	3.7e-02	1.00	0.3935	1.00	Nervous system

Carbonic anhydrase 5A	THROMBOCYTOPENIA	216	77	0.0779	0.0000	inf	2.7e-02	1.00	0.3565	1.00	Blood and lymphatic system
Carbonic anhydrase 5B	THROMBOCYTOPENIA	191	70	0.0857	0.0000	inf	2.0e-02	1.00	0.3665	1.00	Blood and lymphatic system
Carbonic anhydrase 5A	LEUKOPENIA	216	61	0.0984	0.0000	inf	9.9e-03	1.00	0.2824	1.00	Blood and lymphatic system
Carbonic anhydrase 5B	LEUKOPENIA	191	54	0.1111	0.0000	inf	6.4e-03	1.00	0.2827	1.00	Blood and lymphatic system
Dopamine D1 receptor	LEUKOPENIA	374	120	0.0417	0.0000	inf	3.6e-02	0.80	0.3209	1.00	Blood and lymphatic system
Carbonic anhydrase 5B	AGITATION	191	75	0.0800	0.0000	inf	2.6e-02	1.00	0.3927	1.00	Psychiatric
Muscarinic acetylcholine receptor M2	SOMNOLENCE	567	180	0.0389	0.0000	inf	3.4e-02	1.00	0.3175	1.00	Nervous system
Muscarinic acetylcholine receptor M1	TREMOR	355	72	0.0972	0.0035	27.5	7.2e-03	0.88	0.2028	1.00	Nervous system

Muscarinic acetylcholine receptor M5	TREMOR	357	72	0.0972	0.0035	27.7	7.0e-03	0.88	0.2017	1.00	Nervous system
Muscarinic acetylcholine receptor M3	TREMOR	503	96	0.0729	0.0025	29.7	3.7e-03	0.88	0.1909	1.00	Nervous system
Muscarinic acetylcholine receptor M2	TREMOR	567	114	0.0526	0.0022	23.8	3.4e-02	0.86	0.2011	1.00	Nervous system
Serotonin 6 (5-HT6) receptor	TREMOR	546	98	0.0510	0.0045	11.4	4.7e-02	0.71	0.1795	1.00	Nervous system
Muscarinic acetylcholine receptor M4	TREMOR	357	71	0.1127	0.0070	16.1	8.1e-03	0.70	0.1989	0.99	Nervous system
Norepinephrine transporter	TREMOR	508	93	0.0860	0.0096	8.9	1.7e-02	0.67	0.1831	0.99	Nervous system
Serotonin 6 (5-HT6) receptor	CONSTIPATION	546	140	0.0429	0.0025	17.4	4.1e-02	0.86	0.2564	1.00	Gastrointestinal
Muscarinic acetylcholine receptor M5	CONSTIPATION	357	93	0.0645	0.0076	8.5	4.6e-02	0.75	0.2605	0.99	Gastrointestinal
Muscarinic acetylcholine receptor M1	CONSTIPATION	355	92	0.0652	0.0076	8.6	4.3e-02	0.75	0.2592	0.99	Gastrointestinal

Muscarinic acetylcholine receptor M3	CONSTIPATION	503	135	0.0444	0.0054	8.2	4.8e-02	0.75	0.2684	0.99	Gastrointestinal
Carbonic anhydrase 5B	TOXIC EPIDERMAL NECROLYSIS	191	28	0.1786	0.0061	29.1	4.7e-03	0.83	0.1466	0.99	Skin and subcutaneous tissue
Carbonic anhydrase 5A	TOXIC EPIDERMAL NECROLYSIS	216	28	0.1429	0.0106	13.4	3.4e-02	0.67	0.1296	0.99	Skin and subcutaneous tissue
Carbonic anhydrase 2	TOXIC EPIDERMAL NECROLYSIS	558	73	0.0822	0.0062	13.3	2.1e-02	0.67	0.1308	0.99	Skin and subcutaneous tissue
Carbonic anhydrase 5A	AGRANULOCYTOSIS	216	54	0.0926	0.0062	15.0	4.1e-02	0.83	0.2500	0.99	Blood and lymphatic system
Carbonic anhydrase 5B	AGRANULOCYTOSIS	191	48	0.1042	0.0070	14.9	3.0e-02	0.83	0.2513	0.99	Blood and lymphatic system
Carbonic anhydrase 5A	APLASTIC ANAEMIA	216	25	0.2000	0.0052	38.2	5.5e-03	0.83	0.1157	0.99	Blood and lymphatic system
Carbonic anhydrase 5B	APLASTIC ANAEMIA	191	21	0.2381	0.0059	40.5	1.9e-03	0.83	0.1099	0.99	Blood and lymphatic system

Carbonic anhydrase 5B	GLYCOSURIA	191	15	0.3333	0.0057	58.7	6.2e-04	0.83	0.0785	0.99	Renal and urinary
Carbonic anhydrase 5B	JAUNDICE CHOLESTATIC	191	20	0.2500	0.0058	42.8	1.9e-03	0.83	0.1047	0.99	Hepatobiliary
Carbonic anhydrase 5A	JAUNDICE CHOLESTATIC	216	22	0.1818	0.0103	17.6	1.9e-02	0.67	0.1019	0.99	Hepatobiliary
Carbonic anhydrase 5B	PHOTOSENSITIVITY REACTION	191	37	0.1351	0.0065	20.8	1.3e-02	0.83	0.1937	0.99	Skin and subcutaneous tissue
Type-1 angiotensin II receptor	HYPERKALAEMIA	304	24	0.1667	0.0036	46.7	1.7e-02	0.80	0.0789	1.00	Metabolism and nutrition
Muscarinic acetylcholine receptor M3	CONVULSION	503	132	0.0455	0.0054	8.4	4.8e-02	0.75	0.2624	0.99	Nervous system
Muscarinic acetylcholine receptor M2	NASAL CONGESTION	567	40	0.1250	0.0038	32.9	7.5e-03	0.71	0.0705	1.00	Respiratory, thoracic and mediastinal
Serotonin 6 (5-HT6) receptor	DRY MOUTH	546	94	0.0532	0.0044	12.0	4.1e-02	0.71	0.1722	1.00	Gastrointestinal
α-1d adrenergic receptor	OEDEMA	512	127	0.0709	0.0104	6.8	2.6e-02	0.69	0.2480	0.99	General
β-1 adrenergic receptor	FATIGUE	479	133	0.0677	0.0116	5.9	3.4e-02	0.69	0.2777	0.99	General

Carbonic anhydrase 5B	HYPERURICAEMIA	191	11	0.3636	0.0111	32.7	2.1e-03	0.67	0.0576	0.99	Metabolism and nutrition
Carbonic anhydrase 5B	EPIGASTRIC DISCOMFORT	191	9	0.4444	0.0110	40.4	1.8e-03	0.67	0.0471	0.99	Gastrointestinal
Carbonic anhydrase 5B	ORTHOSTATIC HYPOTENSION	191	32	0.1250	0.0126	9.9	4.9e-02	0.67	0.1675	0.99	Vascular
Carbonic anhydrase 5A	PANCREATITIS	216	27	0.1481	0.0106	14.0	3.1e-02	0.67	0.1250	0.99	Gastrointestinal
Carbonic anhydrase 5B	VASCULAR PURPURA	191	22	0.1818	0.0118	15.4	1.8e-02	0.67	0.1152	0.99	Skin and subcutaneous tissue
Carbonic anhydrase 5B	STEVENS-JOHNSON SYNDROME	191	30	0.1333	0.0124	10.7	4.0e-02	0.67	0.1571	0.99	Skin and subcutaneous tissue
Carbonic anhydrase 5B	BLOOD URIC ACID INCREASED	191	13	0.3077	0.0112	27.4	3.6e-03	0.67	0.0681	0.99	Investigations
Carbonic anhydrase 5B	TUBULOINTERSTITIAL NEPHRITIS	191	12	0.3333	0.0112	29.8	2.8e-03	0.67	0.0628	0.99	Renal and urinary